

---

**Estudos dos gastos da União para o  
enfrentamento da COVID-19 durante 2020**

---

**Salomão Oliveira Alves**



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO

Uberlândia  
2021



**Salomão Oliveira Alves**

**Estudos dos gastos da União para o  
enfrentamento da COVID-19 durante 2020**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. André Ricardo Backes

Uberlândia

2021



*Aos meus pais, Maurício Alves de Queiroz e Ivete Maria de Oliveira Alves,  
a meus avós Dalva Maria de Queiroz e Pedro Pinto de Oliveira, Neli Rodrigues de Souza  
e José Maria de Oliveira,  
as minhas irmãs Barbará Dalva Alves de Matos e Mônica Neli Alves,  
aos meus cunhados, Glauber Rezende de Matos e Laudecir Lemos Raiol Júnior,  
e as minhas sobrinhas Isabel Alves de Matos e Amélia Alves de Matos.*



---

# Agradecimentos

Primeiramente, gostaria de agradecer a Deus por todas as oportunidades, capacidades e momentos dados por Ele, desde antes da vida acadêmica quanto durante. Sempre me guinando e iluminando durante essa jornada chamada vida.

Aos meus pais, por todas as lições de vida, experiência e conhecimento passado por eles. Por batalharem com tanto vigor sempre com honestidade para que eu pudesse ter a oportunidade de realizar a graduação. Por sempre entender minhas escolhas e me apoiar incondicionalmente nas minhas decisões, espero sempre honrar esses ensinamentos.

As minhas irmãs, por serem como boas irmãs mais velhas, sempre me ensinado a pensar em todas possibilidades, me corrigindo quando necessário e me ensinado os melhores caminhos. Por sempre estarem ao meu lado, aproveitando e se alegrando com notícias empolgantes e me dando atenção quando necessário desabafar.

Aos meu cunhados, por todos os ensinamentos de vida social e acadêmica. As minhas sobrinhas, por sempre espalharem alegria, risadas e amor pela casa. Aos meus avós, por sempre serem batalhadores e honestos, passando valores justo para as novas gerações. Por sempre nos receber com alegria e felicidades, tornando esses momentos sempre de alegria e diversão.

Aos meu mestres, por todo os ensinamentos teóricos e práticos, além das lições que levarei por toda a vida. Em especial, ao meu orientador Professor Doutor André Ricardo Backes, por me aceitar como orientando e me ajudando sempre quando necessário durante a realização dessa pesquisa.



---

# Resumo

Reconhecida no início de 2020 como pandemia pela Organização Mundial da Saúde (OMS), o surto de COVID-19 causou diversas vítimas ao redor do mundo, mas não apenas pela sua mortalidade, mas também pois as medidas para conter o avanço do vírus prejudica a economia global. Este trabalho busca entender como o Governo Federal do Brasil destinou as verbas públicas e se os estados mais atingidos receberam recursos além de entender a situação geral de cada estado analisado. Esse pesquisa utilizou de técnicas de *Web Scraping*, a linguagem Python e o *data store Big Query* para realizar a extração dos dados, já a análise foi realizada usando a mesma linguagem em um *jupyter notebook* e utilizando-se de gráficos para visualizar os dados. Verificou-se que a situação de alguns estados estava crítica mas que medidas foram tomadas, mas que medidas, por meio das Ações Orçamentárias foram tomadas. A partir desses resultados podemos concluir que, apesar de algumas falhas, a União conseguiu ajudar, de forma geral, os estados analisados no momento de necessidade.

**Palavras-chave:** Pandemia, COVID-19, Economia, Análise de Dados, União.



---

## Lista de ilustrações

|  |    |
|--|----|
| Figura 1 – Raspagem Web . . . . .                                      | 21 |
| Figura 2 – Fluxograma da Gramática dos Gráficos . . . . .              | 26 |
| Figura 3 – Exemplo do Conjunto de Dados das Ações - BigQuery . . . . . | 36 |
| Figura 4 – Dados Extras - BigQuery . . . . .                           | 37 |
| Figura 5 – IDH das Unidades Federativas . . . . .                      | 41 |
| Figura 6 – Histogramas dos Casos de SRAG . . . . .                     | 42 |
| Figura 7 – Evolução das ocorrências por Sexo e Raça . . . . .          | 43 |
| Figura 8 – Ocorrência por estados . . . . .                            | 43 |
| Figura 9 – Distribuição dos Leitos de UTI . . . . .                    | 44 |
| Figura 10 – Ocupação dos Leitos Hospitalares . . . . .                 | 45 |
| Figura 11 – Fornecedores de Respiradores . . . . .                     | 46 |
| Figura 12 – Gastos e Quantidade de Respiradores . . . . .              | 46 |
| Figura 13 – Valor Pago e Faltante das Ações Orçamentárias . . . . .    | 47 |
| Figura 14 – Gastos por Órgão Superior e Fase da Despesa . . . . .      | 48 |
| Figura 15 – Visão Geral da Situação do Brasil - Parte 01 . . . . .     | 51 |
| Figura 16 – Visão Geral da Situação do Brasil - Parte 02 . . . . .     | 52 |



---

## Lista de tabelas

|  |    |
|--|----|
| Tabela 1 – Gastos das Ações Orçamentárias por Estado . . . . . | 50 |
|--|----|



---

## Lista de códigos

|     |  |    |
|-----|--|----|
| 5.1 | Save BigQuery . . . . .                          | 33 |
| 5.2 | Exemplo de código usando Selenium . . . . .      | 34 |
| 5.3 | Exemplo de código usando BeautifulSoup . . . . . | 35 |
| 5.4 | Extração dos Identificadores . . . . .           | 35 |
| 5.5 | Arquitetura Objeto acao_ . . . . .               | 38 |
| 5.6 | Leitura dos Dados . . . . .                      | 38 |
| 5.7 | Função show_info() . . . . .                     | 39 |



---

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUÇÃO . . . . .</b>                           | <b>17</b> |
| 1.1      | Objetivos . . . . .                                   | 18        |
| 1.2      | Organização do trabalho . . . . .                     | 19        |
| <b>2</b> | <b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>                | <b>21</b> |
| 2.1      | Raspagem Web . . . . .                                | 21        |
| 2.2      | Data Munging . . . . .                                | 22        |
| 2.3      | Análise Exploratória dos Dados - AED . . . . .        | 23        |
| 2.4      | Storytelling . . . . .                                | 24        |
| 2.5      | Gramática dos Gráficos . . . . .                      | 24        |
| 2.6      | Visualização dos Dados . . . . .                      | 25        |
| <b>3</b> | <b>TRABALHOS RELACIONADOS . . . . .</b>               | <b>27</b> |
| <b>4</b> | <b>METODOLOGIA . . . . .</b>                          | <b>29</b> |
| 4.1      | Aquisição dos Dados . . . . .                         | 29        |
| 4.2      | Python . . . . .                                      | 29        |
| 4.3      | Ambiente de Desenvolvimento Integrado - IDE . . . . . | 30        |
| 4.4      | Big Query . . . . .                                   | 31        |
| <b>5</b> | <b>EXPERIMENTOS . . . . .</b>                         | <b>33</b> |
| 5.1      | Extração, Transformação e Carregamento . . . . .      | 33        |
| 5.1.1    | Web Scraping . . . . .                                | 33        |
| 5.1.2    | GET Ação . . . . .                                    | 35        |
| 5.1.3    | GET IDH . . . . .                                     | 36        |
| 5.1.4    | GET Leitos de UTI . . . . .                           | 37        |
| 5.1.5    | SET Datasets . . . . .                                | 37        |
| 5.2      | Pré-Análise . . . . .                                 | 37        |
| 5.2.1    | GET Data . . . . .                                    | 38        |

|            |   |           |
|------------|---|-----------|
| 5.2.2      | Data Munging . . . . .                    | 39        |
| <b>5.3</b> | <b>Análises . . . . .</b>                 | <b>40</b> |
| 5.3.1      | Núcleo 1 . . . . .                        | 40        |
| 5.3.2      | Núcleo 2 . . . . .                        | 45        |
| 5.3.3      | Núcleo 3 . . . . .                        | 48        |
| <b>6</b>   | <b>CONCLUSÃO . . . . .</b>                | <b>53</b> |
| <b>6.1</b> | <b>Avaliação dos resultados . . . . .</b> | <b>53</b> |
| <b>6.2</b> | <b>Avaliação futuras . . . . .</b>        | <b>54</b> |
|            | <b>REFERÊNCIAS . . . . .</b>              | <b>55</b> |

---

## Introdução

Segundo Larsson (2001), a definição para globalização é “o processo de encolhimento do mundo, de encurtamento das distâncias, coisas se aproximando. Diz respeito à facilidade crescente com que alguém de um lado do mundo pode interagir, para mútuo benefício, com alguém do outro lado do mundo”. Então, segundo essa definição, o mundo se tornou menor, uma *aldeia global*, onde fatos ocorridos em um determinado local, rapidamente, refletem em várias partes do mundo. Essa proximidade facilitou a troca de informação, cultural, de conhecimento e afins, porém, doenças também começaram a se espalhar, sendo uma pandemia classificada, por Bill Gates, como uma das três principais ameaças à humanidade (FENASAUDE, 2021).

Como previsto por Gates, no final do ano de 2019, mais precisamente no mês de dezembro, na província de *Wuhan*, na China (ISTOEDINHEIRO, 2021), começaram a surgir os primeiros casos de uma doença que passaria a ser conhecida como COVID-19, causada pelo vírus *SARS-CoV-2*. Demorou apenas alguns meses para a Organização Mundial de Saúde (OMS) decretá-la como uma pandemia e, a partir desse momento, nossas vidas mudaram bastante (BCCNEWS, 2021a). Apesar de ter passado mais de um ano, pouco se sabe sobre a origem desse vírus. No relatório mais recente da OMS sobre a origem do vírus não se chega a uma conclusão definitiva, apenas que é provável que o vírus tenha vindo de um animal e tenha passado por um hospedeiro intermediário, e sendo improvável que seja resultado de um acidente de laboratório (BCCNEWS, 2021b).

Como já é sabido, apesar do Brasil possuir um dos maiores PIBs, porém, em contra partida, ele também possui altos índices de desigualdade social. Com o anúncio de uma pandemia, diversos centros populacionais decretaram *lockdown*, para diminuir o contágio, e diversas medidas públicas foram adotadas para combater essa crise sanitária. Devido ao isolamento social, um país que já era desigual se tornou cada vez mais, muitos serviços em que a interação era extremamente necessárias, como garçons, academias, pequenos restaurantes, dentre outros, foram prejudicados, passando meses sem renda, agravando ainda mais a situação já complicada do país.

Diante de cenário devastador, com uma crise sanitária já instalada e causando muitos

óbitos, além de uma crise econômica que, conseqüentemente, iria atingir a população, principalmente as de classes mais baixas, o Governo Federal precisou agir para tentar diminuir o impacto que as medidas de distanciamento social causariam na economia e ao mesmo tempo fornecer suporte e logística para tratar os doentes. Para isso, o governo, por meio de Ações Orçamentárias, disponibilizou verbas para diferentes órgãos e estados, seja para combater a crise sanitária, com compras de equipamentos de proteção e médicos, ou a crise econômica, devido ao fechamento dos comércios.

## 1.1 Objetivos

Com uma crise dessa magnitude, os impactos causados na sociedade são grandes, por isso, o objetivo dessa pesquisa é analisar as ações orçamentárias para entender aonde o dinheiro foi enviado (seja para qual estado ou órgão/ministério) e qual foi a finalidade dele.

Além das ações orçamentárias, também dados sobre o Índice de Desenvolvimento Humano (IDH) de cada estado, assim como o histórico da quantidade de leitos de Unidades de Terapia Intensiva (UTI) e a de casos e óbitos da Síndrome Respiratória Aguda Grave (SRAG) de cada estado, além da quantidade de respiradores, sendo os três últimos e as ações orçamentárias referenciando-se ao período de Fevereiro de 2020 até Dezembro de 2020 e o IDH referente ao Censo Demográfico de 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Durante a análise, foram criados três núcleos principais de análises. Chamaremos eles de:

- ❑ Núcleo 1: *Overview* do Brasil durante a pandemia, em que foram usados todos os conjuntos, menos os das ações;
- ❑ Núcleo 2: *Overview* das Ações Orçamentárias, em que foram utilizados o conjunto das ações;
- ❑ Núcleo 3: Relação entre a situação do Brasil e as ações orçamentárias.

O primeiro núcleo, como já dito, utilizou os seguintes conjuntos: o IDH, leitos hospitalares, casos e óbitos de SRAG, ocupação dos leitos e quantidade de respiradores. O objetivo por trás desse núcleo é obter um vislumbre melhor da situação em que cada unidade federativa estava em cada conjunto de dados. No final, uma compilação dos principais dados que será utilizado no último núcleo.

O próximo núcleo usou apenas o conjunto de dados das ações, para se ter uma noção de como a União destinou as verbas. Nele, foi apresentada uma visualização geral dos dados das Ações Orçamentárias, em outras palavras, apenas a *big picture* dos dados,

selecionando as principais. Depois, um compilado dos dados foi montado para a análise no próximo núcleo.

Por último, utilizou os dois conjuntos de dados gerados nos núcleos acima. Assim, temos os conhecimentos que foram extraídos no primeiro núcleo, ou seja, o entendimento da situação do Brasil durante a pandemia, e, a noção dos gastos da União. Com esses dois núcleos de informações, é possível analisar de forma mais precisa aonde o Estado injetou verba pública e se aquele lugar realmente precisava, tanto quanto os lugares necessitados e que não receberam ajuda.

## 1.2 Organização do trabalho

O trabalho se divide em seis capítulos, sendo elas Introdução, Fundamentação Teórica, Trabalhos Relacionados, Metodologia, Experimentos e Conclusão.

A Introdução busca motivar e contextualizar o leitor sobre o trabalho de acordo com a literatura, bem como mostrar os objetivos e organização do trabalho. A Fundamentação Teórica busca exibir ao leitor os fundamentos e conceitos necessário para compreender a pesquisa. Já os Trabalhos Relacionados, buscam exibir o que esta acontecendo na área de estudo.

No capítulo Metodologia, são apresentadas as técnicas, ferramentas e soluções utilizadas para realizar a extração dos dados, seu pré processamento e a análise. No capítulo Experimentos, são apresentadas as análises realizadas e seus resultados. Por último, o capítulo Conclusão apresenta a avaliação dos resultados obtidos, e discute possíveis melhorias e investigações futuras.



---

## Fundamentação Teórica

O capítulo irá apresentar os principais conceitos que serão utilizados durante a pesquisa, apresentando a explicação que a literatura traz sobre eles.

### 2.1 Raspagem Web

*Web Scraping* ou Raspagem Web (Figura 1) é um processo no qual um *bot* irá extrair conteúdos e dados de um determinado *website*. São utilizados processos automáticos, no caso esses seriam os *bots*, comumente chamados de robôs. Eles realizam a “raspagem” de informação, que seria o ato de acessar o código escrito em *HyperText Markup Language* (HTML), que contém todas as informações mostradas na página, selecionando as desejadas e salvando-as em algum banco de dados ou então em algum arquivo, para posterior recuperação ou análise.



Figura 1 – Raspagem Web

Existem muitos seguimentos diferentes em que pode se aplicar técnicas de *Web Scraping*, Saurkar, Pathare e Gode (2018) escrevem sobre algumas, como a detecção de alterações no site, comparação do produto, transmissão do tempo e monitoramento de dados, análise de pesquisa, análise de dados em gráficos, indexação da web e verificação de classificação, análise de publicidade e de mercado.

Ainda em (SAURKAR; PATHARE; GODE, 2018), os autores realizam uma análise geral sobre as principais técnicas de raspagem, são elas:

1. **Clássico Cópia e Cola** onde uma pessoa, de forma manual, copia e cola todas as informações desejadas;
2. **Programação de *HyperText Transfer Protocol* (HTTP)** utilizando-se de programação de *sockets*, o usuário consegue recuperar dados ao fazer solicitações HTTP ao servidor.
3. **Análise em *HyperText Markup Language* (HTML)** utiliza-se de ferramentas para acessar as páginas HTML (neste trabalho utilizou-se da linguagem Python e suas bibliotecas Selenium e BeautifulSoup);
4. **Análise do *Document Object Model* (DOM)**, isto é, a análise da árvore DOM extraída das páginas web;
5. **Software de Raspagem** são soluções construídas exatamente para a realização dessa tarefa de raspagem;
6. **Analísadores de Páginas da Web com Visão Computacional** são algoritmos de *Machine Learning* e visão computacional criados para identificar e extrair informações.

## 2.2 Data Munging

No mundo da Ciência de Dados há dois grandes problemas enfrentados por cientistas e empresas, o primeiro é a grande quantidade de dados semi ou não estruturados, sem rótulos, desorganizados, e o outro seria a falta de dados para realizar determinadas tarefas. Nessa seção é apresentado sobre o trabalho pesado de limpar e organizar os dados (SKIENA, 2017).

A parte responsável pela limpeza e organização dos dados, também conhecida como *Data Wrangling* ou *Data Integration*, é de extrema importância, consumindo de 50 a 80 por cento do tempo de um cientista de dados (LENZERINI, 2018).

O motivo dessa tamanha importância é simples: um modelo de inteligência artificial ou aprendizado de máquina, até mesmo uma simples visualização de dados, todas essas aplicações irão apenas refletir o que os dados estiverem apresentando. Portanto, se os dados estiverem mal organizados, forem “lixo”, então a aplicação irá gerar “lixo”. Por isso, tópicos como *Data Quality* são tão comentados, pois verificar a qualidade dos dados e transformá-los em uma forma mais amigável para as aplicações é importante.

Para (ENDEL; PIRINGER, 2015), praticamente todo conjunto de dados, quando analisado pela primeira vez, apresenta desafios. O autor os dividiu nas seguintes categorias:

1. **Recursos Básicos:** devem ser explorados os aspectos básicos, como tamanho, formato e afins;
2. **Qualidade dos Dados:** de extrema importância, existindo diversas métricas na literatura para comprová-la;
3. **Mesclar e Vincular:** muitas vezes é necessário conectar ou mesclar dois ou mais conjuntos de dados;
4. **Reprodutibilidade e Documentação:** é a capacidade de reproduzir os dados, com o auxílio da documentação dos mesmos;
5. **Big Data:** é a quantidade gigantesca de dados que é produzida ao redor do mundo;
6. **Dirty Data:** são criadas estratégias para trabalhar com a sujeira que pode ser encontrada nos dados;
7. **Incerteza dos Dados:** há diversas técnicas na literatura para medir a incerteza que os dados apresentam;
8. **Tolerância a Erros:** a capacidade da aplicação conseguir tolerar erros, esperados ou não, e sujeiras nos dados;
9. **Transformação e Edição:** transformam um conglomerado de dados em um conjunto estruturado e útil.

## 2.3 Análise Exploratória dos Dados - AED

O principal motivo de se realizar uma análise é observar o que os dados de um determinado conjunto de dados estão querendo nos dizer, verificar tendências e padrões, além de possíveis incoerências, podendo assim formular novas hipóteses e teorias. Sendo definida por Tukey (1962) como: “Procedimentos para analisar dados, técnicas para interpretar os resultados de tais procedimentos, formas de planejar a reunião dos dados para tornar sua análise mais fácil, mais precisa ou mais exata e toda a maquinaria e os resultados da estatística (matemática) que se aplicam à análise de dados.”

Para MEDRI (2011), uma boa análise irá extrair a maior quantidade possível de informação que um conjunto de dados tem, onde serão utilizadas posteriormente para tomadas de decisões ou para alimentar algum modelo de *Machine Learning* ou alguma aplicação de visualização.

Ainda segundo MEDRI (2011), a análise conterà diversos métodos exploratórios e estatísticos, os quais estão passando por uma evolução, se tornando mais robustos, complexos e com melhor desempenho, para assim, realizarem o estudo detalhado do dados, antes de adaptá-los para uma determinada aplicação. Como o próprio Tukey (1962) diz:

“A análise de dados é um campo maior e mais variado do que inferência, procedimentos incisivos ou alocação.”

Como Behrens (1997) sugere, a AED nos diz que, para diferentes estágios da pesquisa, irão existir diferentes níveis de especificidade da hipótese ou da conclusão, e de tipos de perguntas também. Sendo vista como indispensável em qualquer investigação científica, “A análise exploratória de dados nunca pode ser a história toda, mas nada mais pode servir como a pedra fundamental - como primeiro passo” (TUKEY, 1962).

Com o avanço e o surgimentos de novos métodos e técnicas, com resultados cada vez melhores, começaram a surgir diversas linguagens de computação ou até mesmo pacotes para determinada linguagem, os quais eram focados na parte estatística. Como, por exemplo, a Linguagem R, que foi construída com base na Linguagem S da *Bell Laboratories* (COOK; SWAYNE, 2007).

## 2.4 Storytelling

Desde os tempos das cavernas o ser humano é um contador de histórias. Seja para se vangloriar sobre suas conquistas nas caças ou para passar tradições e experiência. *Storytelling* é simplesmente a capacidade de desenvolver boa narrativa, ser um contador de histórias e, assim, a informação é transmitida por meio dessa história, envolvendo os ouvintes e fixando mais em quem escuta. Esta técnica é utilizada quando se tem o objetivo de transmitir o que foi analisado em um conjunto de dados.

Por mais que o tema seja desgastado, o fato de ser criada uma história, um material único, ele ainda vai se destacar por abranger a perspectiva do autor, levando o público para uma jornada, além de gerar emoções e identificações. Como dizem Gershon e Page (2001), “Uma boa história transmite um grande quantidade de informação em relativamente poucas palavras em um formato que é facilmente assimilado pelo público”.

Mas fica a pergunta, o que torna o *Storytelling* tão valioso? Segundo Gershon e Page (2001), “meio visuais (...) são normalmente usados para resolver um problema ou transmitir informações científicas específicas”. E continuam dizendo que “ambientes nos quais funções de visualização de informações envolvem massivos fluxos de informações e fontes de dados que chegam em tempo real ou de dados existentes”.

## 2.5 Gramática dos Gráficos

A Gramática dos Gráficos (GOG na sigla em inglês) é um sistema para padronizar a forma como se deve criar um gráfico, sendo composto por sete camadas (ou componentes) ortogonais (WILKINSON, 2012). O fato de serem ortogonais significa que cada componente é um conjunto de elementos, os quais são combinados e cada elemento escolhido irá influenciar significativamente no resultado final. Portanto, é basicamente um sistema

combinatório de componentes e blocos, possuindo um alto grau de expressividade, ou seja, produz uma grande variedade de gráficos.

Uma gramática bem estruturada permite entender como o gráfico foi construído (a forma como os componentes foram misturados), podendo revelar conexões entre gráficos visualmente diferentes entre si. Assim, o entendimento da fundação da construção de qualquer gráfico é forte, ajudando a formar gráficos corretos e bem montados, apesar de poder existir gráficos gramaticalmente corretos mas que não fazem sentido. Segundo Wickham (2010), uma boa gramática é apenas o primeiro passo para criar uma boa sentença.

Com a descrição desse sistema em mente, percebemos que ele é muito mais do que apenas uma taxonomia, é uma descrição de o que é a construção de gráficos estatísticos. Nas palavras do Wilkinson (2012): “É um sistema computacional baseado na matemática clássica de representar funções e relações em espaços cartesianos e outros”.

A Figura 2 mostra o diagrama contendo os setes componentes da gramática (WILKINSON, 2012). A primeira camada é a das **Variáveis**, ela é o alicerce de todos os gráficos, nele iremos decidir quais variáveis irão para o gráfico e em qual eixo vão pertencer. Depois vem a **Álgebra**, na qual são realizadas combinações entre as variáveis. Já na terceira camada **Escala** é onde começamos a configurar a sua estrutura, adicionando as devidas escalas.

A camada **Estatística** é responsável por realizar a análise estatística dos dados e apresentar os resultados de forma resumida. A próxima camada, **Geometria**, sendo responsável por definir ou adicionar a posição, intervalos, forma, tamanho, linhas, basicamente é o desenho do gráfico. A penúltima camada, **Coordenadas**, passa as informações da área na qual o gráfico será construído. A última camada, **Estética**, é o resultado final.

## 2.6 Visualização dos Dados

Segundo Aparicio e Costa (2015), temos uma mente muito visual, ou seja, nos damos bem com elementos visuais. Visualização de Dados é uma necessidade antiga, diversas línguas visuais foram criadas (*Phaistos*, *Sumerian* e *Assyrian cuneiform*) baseando-se mais no visual do que no som. Essa habilidade de transmitir informações por meios de tabelas, gráficos e mapas (elementos visuais) é a definição de Visualização de Dados.

Para Post, Nielson e Bonneau (2002), o surgimento do campo Visualização de Dados iniciou nos primórdios da computação, na década de 50. Porém, passou a ganhar maior relevância após um relatório da *National Science Foundation* (NSF) demonstrando a importância do campo. A Visualização de Dados é a junção de outros dois métodos, Visualização Científica e Visualização de Informação, o primeiro é relacionado as imagens geradas durante um processo de computação científica e o segundo para visualizar informações de modo geral. (POST; NIELSON; BONNEAU, 2002).

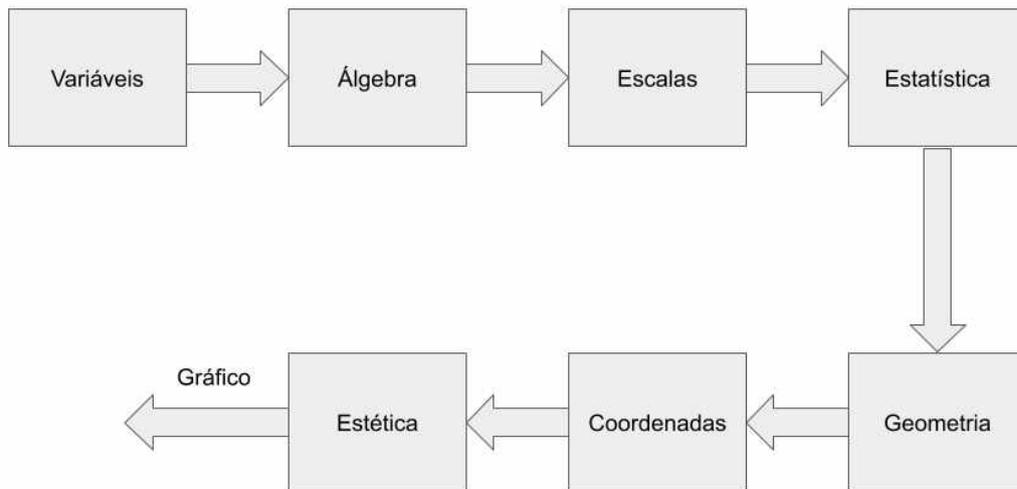


Figura 2 – Fluxograma da Gramática dos Gráficos

Como toda técnica, são necessárias regras, por isso da Gramática dos Gráficos. Adicionando uma narrativa, o *Storytelling*, cria-se uma maneira mais acessível de ver e entender discrepância, tendência e padrões nos dados. Sendo utilizadas em diversas áreas, desde observações empíricas até cartografia temática (FRIENDLY, 2008).

---

## Trabalhos Relacionados

Os autores Albertini, Backes e SÁ (2019) analisaram as principais características dos pesquisadores brasileiros na área de Ciência da Computação. Utilizando-se dos dados coletados do Ministério da Ciência e Tecnologia e do Ministério da Educação, foram analisados diversos aspectos da carreira dos pesquisadores, como a trajetória de carreira e produção científica. Baseando sua análise em Way et al. (2017), o objetivo é descobrir a taxa de produção de artigos científicos de acadêmicos e se seguia a regra convencional, de declínio gradual de aumento rápido e pico curto. Porém, infelizmente, não se pode realizar o mesmo experimento, pois muitos padrões definidos por Way et al. (2017) seriam inviáveis de serem aplicados para os dados brasileiro devido à pequena quantidade de informação disponível. Com o auxílio de diversos gráficos de linhas e pontos, um *HeadMap*, tabelas e de duas funções lineares. No final do artigo, os autores chegam à mesma conclusão que Way et al. (2017), que somente para algumas trajetórias a narrativa convencional sobre reprodução de artigo é vista. Além de terem demonstrado, ao decorrer do artigo, que a academia brasileira de Ciência da Computação ainda é subdesenvolvida.

Em Oliveira, Neto e Ferreira (2006), é monitorada a quantidade de elementos metálicos em combustíveis, uma vez que esses elementos são prejudiciais aos veículos, realizando esse monitoramento, como eles explicam em: “por meio de um controle analítico eficiente dos teores de aditivos inorgânicos incidentais ou acidentais, é possível gerenciar, (...) a qualidade dos combustíveis oferecidos ao consumidor“. Como Oliveira, Neto e Ferreira (2006) cita: “Em diversos segmentos da química o crescimento do uso de recursos computacionais para analisar dados químicos cresceu nos últimos anos”, por isso é utilizado a Análise de Componentes Principais (PCA), um método de análise multivariada usado para projetar conjuntos de dados de n-dimensões com menos dimensões, duas ou três. Para desenvolver a pesquisa, os autores utilizaram-se de diversas fórmulas relacionadas à química, o já citado PCA e diversos gráficos. Os autores concluíram que a utilização do PCA foi de grande valia para a extração de informações relevantes.

Já em Neves et al. (2000), é estudado como os avanços na capacidade de processamento computacional e os sistemas de informação geográfica (SIG) evoluíram. Além de que

algumas técnicas estatísticas foram adaptadas para melhor trabalhar com dados espaciais, aumentando o leque de opções que podem ser usadas para melhor realizar a análise espacial, e que, nas palavras de Neves et al. (2000), “Este conjunto de ferramentas é definido na literatura como sendo uma coleção de técnicas para descrever e visualizar distribuições espaciais, identificar situações atípicas, descobrir padrões (...)”. O conjunto de ferramentas escolhido foi o SPRING (Sistema para Processamento de Informações Georeferenciadas). Para comprovar a usabilidade dessas técnicas, o autor utiliza dados sócio-econômicos oriundos do mapeamento da exclusão social da cidade de São Paulo. Para conduzir o trabalho, além de diversas fórmulas e equações definidas, os autores se utiliza de gráficos de pontos e linhas, além de gráficos de mapas (da cidade de São Paulo) e de tabelas. Ao final, os autores concluem que o conjunto de ferramentas foi de grande valia, como eles citam em: “foi possível verificar que este conjunto de ferramentas aumenta a possibilidade de compreensão da dinâmica espacial dos dados”.

Nos últimos anos, internações por causas sensíveis à atenção ambulatorial (CSSA), doenças que se tratam de forma correta não evoluiriam para um quadro de internação, se tornaram dados interessantes para avaliar a qualidade da atenção primária à saúde. Por isso, Perpetuo e Wong (2006) realizam um estudo combinando esse conjunto de dados com os dados das causas de internação no Sistema de Internações Hospitalares do SUS (SIH-SUS). Para realizar as análises, os autores utilizam de fórmulas matemáticas, com os coeficientes de internação hospitalar (CIH) e alguns outros dados sobre o SUS para complementar a análise, além de diversos gráficos de pontos, linhas e tabelas. No final, os autores ficaram satisfeitos que a análise se mostrou útil para avaliar a qualidade da atenção primária à saúde, apesar de ressaltarem que esses dados devem ser encarados com cautela e uma análise mais profunda é necessária.

Com base nesses cinco trabalhos, podemos observar algumas similaridades e convergências entre eles. Começando pelas diferenças, claramente é percebido que cada uma das pesquisas focam em alguma área ou problema específico, o que demonstra uma grande vantagem da análise de conjuntos de dados, o leque de opções para desenvolver projetos é grande, podendo ser sobre química (OLIVEIRA; NETO; FERREIRA, 2006) ou até mesmo sobre medicina (PERPETUO; WONG, 2006). Sobre as semelhanças, em todos os trabalhos os autores utilizaram de fórmulas matemática e químicas (devidamente explicadas) para auxiliar na análise, já para demonstrar os resultados, tabelas e diversos tipos de gráficos, incluindo *HeadMaps* e Mapas, foram explorados.

---

# Metodologia

Esse capítulo irá apresentar quais técnicas, métodos, soluções, software ou tecnologia foram utilizados para realizar esse trabalho.

## 4.1 Aquisição dos Dados

Para esse trabalho lidamos com sete fontes distintas de dados. Para coletar os números das ações orçamentárias, utilizamos o site do Tesouro Transparente (ECONOMIA, 2021), onde foi possível acessar o arquivo *Comma-Separated Values* (CSV) que alimenta a visualização dos gastos de COVID disponível. Como esse *dataset* continha apenas uma parte das informações sobre as ações, apenas o número delas foi extraído.

Para extrair as ações, acessamos, com o auxílio do Python, a página do Portal de Transparência (UNIAO, 2021), e com os números extraídos da base anterior, foi possível filtrar e utilizar técnicas de *Web Scraping* (Seção 5.1.1), para coletar todas as informações.

A quantidade de leitos de UTI disponíveis durante o período analisado foi retirada do site do Cadastro Nacional de Estabelecimentos de Saúde (SAUDE, 2021a). Já o IDH foi retirado do site do IBGE (ESTATÍSTICA, 2021). Ambas as extrações foram realizadas utilizando as mesmas técnicas de *Web Scraping* descritas na Seção 5.1.1.

Por último, a quantidade de casos e óbitos de SRAG e a de respiradores comprados, além da ocupação de leitos foram extraídos do site *Open Data SUS* (SAUDE, 2021b). Os dados foram obtidos a partir do download do CSV disponível. É importante frisar que, após a extração, alguns desses conjuntos de dados precisaram passar por uma transformação, para melhor entendimento das informações consultar a Seção 5.1.5.

## 4.2 Python

Para executar as técnicas de *Web Scraping* e as análises de dados, foi utilizada a linguagem de programação de alto nível Python (versão 3.7.7), lançada no início da década de 90 pelo holandês Guido Van Rossum. Ela foi escolhida por possuir uma sintaxe mais

simples e próxima da humana, além da quantidade significativa de pacotes disponíveis, principalmente para aplicação de análise estatística. Com o auxílio das próprias funções e estruturas que a linguagem já oferece, utilizamos algumas bibliotecas para realizar tarefas específicas, abaixo passarei um breve resumo sobre todas (TECNOBLOG, 2021).

O *pandas* (versão 1.1.5) e o *numpy* (versão 1.15.4) são duas das bibliotecas utilizadas, as quais são as mais comuns para realizar análise de dados. No caso da primeira, principalmente por oferecer estruturas, como os *dataframes*, e ferramentas de análise de alto nível, facilitando a manipulação dos dados. Já na segunda, por fornecer computação matemática rápida em *arrays* e matrizes (CLOUDXLAB, 2021).

Já para realizar as técnicas de *Web Scraping*, seriam a *Selenium* (versão 3.141.0) para realizar toda a movimentação e interação necessária no *browser*, ou seja, na janela de navegação. E a *Beautiful Soup* (versão 4.9.0) para realizar a pesquisa no código HTML da página que estava sendo analisada pela biblioteca anterior. Já a *time*, a *pytz* e a *datetime* são bibliotecas focadas para a manipulação de datas e fusos horários.

Para a visualização foi usado a *Matplotlib* (versão 3.2.2), ótima biblioteca python para visualizações, oferecendo uma interface orientada a objetos para produzir gráficos. Também será usado a *Seaborn* (versão 0.11.1), que é baseada na anterior, porém oferece uma interface de alto nível para construção de gráficos estatísticos atrativos e informativos.

### 4.3 Ambiente de Desenvolvimento Integrado - IDE

Para a realização dessa pesquisa foram utilizadas três IDE. A primeira ferramenta foi o *Microsoft Visual Studio Code*, desenvolvido pela própria *Microsoft*, a segunda foi o *Jupyter Lab*, desenvolvida pelo Projeto Jupyter e, por último, o *Google Colaboratory*, desenvolvida pela Google.

O *Microsoft Visual Studio*, além do editor e do depurador padrão, contém compiladores, ferramentas de preenchimento de código, designers gráficos e outros recursos para realizar um desenvolvimento de software. Existem muitos recursos disponíveis, como por exemplo, rabiscos, ações rápidas, limpeza de código, refatoração, dentre outros (MICROSOFT, 2021).

Já o *Jupyter Lab* é uma evolução do Jupyter Notebook, uma ferramenta muito popular na área de *Data Science* e *Machine Learning* pois ela permite produzir narrativas computacionais reproduzíveis, uma ótima opção para realizar o *storytelling* (Seção 2.4). A possibilidade de combinar códigos com texto dissertativo, gráficos e tabelas dinâmicas abre um grande leque para melhor apresentar os dados (JUPYTER, 2021).

Por último, temos o *Google Colaboratory*, que é um ambiente colaborativo de *notebooks Jupyter* que não requer configuração e é executado na nuvem. Ele permite a abertura de arquivos *jupyter* externos, além de poder escrever códigos e textos em *markdown*. É um

serviço gratuito oferecido pela Google com o intuito de incentivar a pesquisa nas áreas de *Machine Learning* e *Artificial Intelligence* (ALURA, 2021).

## 4.4 Big Query

Como já dito na Seção 4.1, para alguns conjuntos de dados foi necessária a utilização de técnicas de *Web Scraping* (Seção 2.1) e o *data storage* selecionado foi o *Big Query*, serviço de armazenamento de dados oferecido pelo Google por meio da sua plataforma de serviços em nuvem *Google Cloud Platform* (GCP).

Essa solução permite resolver problemas por meio de consultas SQL super-rápidas com a infraestrutura de processamento da Google. É possível acessá-lo pelo *Console* do *Google Cloud*, ou utilizando alguma linguagem de programação, como Java, .NET ou Python (utilizada na pesquisa) (GOOGLE, 2021). Ele nos permite separar os dados em diversos conjuntos que, por sua vez, possuem diversas tabelas, possibilitando deixar os dados organizados.



---

# Experimentos

Nesse capítulo, será mostrado como os experimentos foram executados. Primeiramente será discutida a forma como os dados foram extraídos e as modificações necessárias feitas neles. Depois, é descrito como foi realizado o carregamento dos dados e o pré-processamento dos mesmos no *script* antes da análise. E, por último, cada núcleo de análise realizado, mostrando as ferramentas e técnicas usadas. Os códigos discutidos na primeira seção foram executados em uma máquina Ubuntu 18.04.4 LTS, 64 bits, processador Intel Core i5-5200U, já os códigos das outras seções utilizaram o *Google Colaboratory*.

## 5.1 Extração, Transformação e Carregamento

Nesta seção, serão apresentadas as técnicas e ferramentas utilizadas para a realização da raspagem e as configurações feitas em alguns conjuntos de dados. Como visto na Seção 4.1, foram selecionados sete conjuntos de dados. Os referentes a casos e óbitos de SRAG, aos respiradores e às ocupações de leitos foram obtidos a partir do *download* de um arquivo CSV.

### 5.1.1 Web Scraping

Para a aplicação dessas técnicas foi utilizada a linguagem de programação *Python* e duas bibliotecas que realizaram as principais funções, *Selenium* e *Beautiful Soup*. Quando utilizada, foi usada a função apresentada no Código 5.1 para salvar os dados no *BigQuery*, o método `to_gbq()` do pacote *Pandas* é utilizado para auxiliar na inserção (credenciais específicas são necessárias para conectar com o GCP). Para tratamento de erros, a mesma função de salvamento é usada.

```
def save_bq(df, dataset, tabela):
    credentials = get_credentials()
    df.to_gbq(destination_table=dataset+'.'+tabela,
              project_id=PROJECT_ID,
              credentials=credentials,
```

```
if_exists='append')
```

Código 5.1 – Save BigQuery

### 5.1.1.1 Selenium

Essa biblioteca permite controlar um *browser* como se fosse um ser humano usando o teclado e mouse, ou seja, permite realizar qualquer ação manualmente possível. Esse tipo de técnica é utilizada quando processos automáticos são necessários. Ela possui a capacidade de acessar diversos *browsers*, como *Firefox*, *Chrome* e *Edge*.

Como pode ser visto no Código 5.2, as primeiras linhas se referem às funções utilizadas do *Selenium*. A função *Chrome()* é usada para realizar a conexão com o navegador (o objeto *browser* é criado), a *Options()* recebe argumentos que são passados para o *driver* selecionado, possibilitando que o navegador não abra (*'-headless'*) ou então que não saia nenhum som (*'-mute-audio'*). Realizamos a requisição com o método *get()*, oriundo do *driver*, passando a URL como parâmetro.

Para verificar se a página está carregando corretamente, foi usada a função *WebDriverWait()*, esperando trinta segundos para que a *tag* com ID passado, utilizando ainda as funções *EC()*, para conseguir visualizar o elemento, a *By()* para definir que é o parâmetro ID do elemento e a *TimeoutException* para tratar quando a página não é carregada. Para tratar as mudanças de páginas, as funções *find\_element\_by\_id()* e *find\_element\_by\_tag\_name()*, também oriundo do *driver*, foram usadas para achar o elemento com o ID passado e executar a ação *click()*, clicando-o e abrindo a próxima página.

```
from selenium.webdriver import Chrome
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.options import Options
from selenium.common.exceptions import TimeoutException
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC

options = Options()
options.add_argument('--headless')
browser = Chrome('./chromedriver', options=options)
browser.get(URL)
try:
    WebDriverWait(browser, 30).until(
        EC.visibility_of_element_located((By.ID, "id_name")))
except TimeoutException:
    pass # trata erro - Pagina nao carregada
try:
    browser.find_element_by_id("id_name").find_element_by_tag_name('tag').click()
except Exception as e:
```

```
pass # tratar erro - Mudanca de Pagina
```

Código 5.2 – Exemplo de código usando Selenium

### 5.1.1.2 BeautifulSoup

Essa biblioteca é feita para a análise de documentos HTML, como foi usado na pesquisa, e XML (*Extensible Markup Language*). Para isso, recebendo como parâmetro o código a ser analisado, ela cria uma árvore para essas páginas, facilitando a pesquisa dos dados.

O código HTML é extraído do *driver*, objeto gerado anteriormente. Depois, a função *BeautifulSoup()* recebe essa informação e transforma em uma “sopa“ (daí vem o nome). Com essa “sopa”, podemos, com o auxílio de métodos como *find()* e *find\_all()*, selecionar as *tags* que o desenvolvedor ache interessante e extrair seu conteúdo.

```
from bs4 import BeautifulSoup

html = browser.page_source
soup = BeautifulSoup(html, "html.parser")
content_rows = soup.find("tbody").find_all("tr")
```

Código 5.3 – Exemplo de código usando BeautifulSoup

### 5.1.2 GET Ação

Para a realização da extração dos dados referentes às Ações Orçamentárias, foi necessário extrair, primeiramente, o código identificador de cada ação. Para isso, no site do Tesouro Transparente (ECONOMIA, 2021), existe uma visualização sobre os gastos para combater a COVID-19. Ele é alimentado por um CSV que está disponível para *download*. Porém, como apenas algumas informações estavam disponíveis, foram extraídos apenas os identificadores por meio da função apresentada no Código 5.4.

```
def get_acao():
    acao_listas=[]
    titulo='Acao Governo Codigo'
    for i in range(11):
        df=pd.read_excel('../PainelCOVIDBase.xlsx', usecols="B", sheet_name=i)
        lista_acoes=df.iloc[:,0].astype(str).unique()
        lista_acoes=[l for l in lista_acoes if l!=titulo and l!='nan']
        acao_listas.extend(lista_acoes)
    acao_df=pd.DataFrame({'Acao': acao_listas})
    return acao_df.iloc[:,0].unique()
```

Código 5.4 – Extração dos Identificadores

Para acessar as informações completas de cada ação, foi acessada a página do Portal de Transparência (UNIAO, 2021). Como a página permite que se filtre apenas as ações desejadas em um período de tempo específico por meio da *Uniform Resource Locator* (URL), as técnicas apresentadas na Seção 5.1.1 foram de grande valia, uma vez que as requisições são feitas diretamente pelas URLs. Ao acessar as páginas, as ações eram formadas por vários programas e, para obter os detalhes desses programas, foi necessário extrair o número desse programa (identificador) e salvá-lo para posterior extração mais detalhada do mesmo.

Para armazenar todos esses dados foi utilizado o *BigQuery*. Os conjuntos de dados foram separados por ações (o nome escolhido foi `acao_IDAcao`), onde cada um deles contém uma tabela chamada de `level01`, a qual continua os dados mais gerais dos programas. Além dessa tabela, eles ainda continuam uma ou mais tabelas, onde cada uma delas representa um programa específico daquela ação (o nome escolhido foi `IDPrograma_detail`) (Veja Figura 3).



Figura 3 – Exemplo do Conjunto de Dados das Ações - BigQuery

Como foi realizada uma extração automática dos dados, as páginas do Portal de Transparência nem sempre carregavam e algumas vezes a própria conexão de *internet* falhava. Como as requisições eram feitas por mês e por ação, esse erro foi tratado para que caso ocorresse algum, eram salvos, em um conjunto de dados no *BigQuery* chamado de *err*, a data, a ação, o programa (caso necessário), a URL, a data do erro e o devido erro, para posterior extração. Também era tratado, na mesma lógica, a mudança de páginas, pois em algumas ações, apareciam várias páginas com conteúdo, para retirar todos os dados, as interações com o *browser* foram usadas.

### 5.1.3 GET IDH

Para extrair os dados do IDH de cada estado brasileiro foi acessado o site do IBGE e utilizando-se da já descrita técnica de *Web Scraping*. O site permite fazer filtros pela URL, mudando as siglas dos estados, sempre em minúsculo, facilitando a sua extração. A mesma lógica para tratamento de erro usada na Seção 5.1.2 para tratar quando a página não é carregada corretamente, é usada neste ponto, salvando os dados.

Em relação à questão de armazenamento de dados (ver Figura 4), também foi utilizado o *BigQuery*. Na questão do tratamento de erros, ele é salvo no conjunto de dados *err* na tabela *idh*. Quando não ocorrem erros, ele armazena no conjunto chamado *dados\_extras* na tabela, também chamada de, *idh*.

#### 5.1.4 GET Leitos de UTI

Para esse conjunto de dados, as técnicas de *Web Scraping* foram novamente utilizadas. Novamente, o site permitia realizar os filtros pela URL, porém, para filtrar os estados, foi necessário, de forma manual, coletar todos esses identificadores e inseri-os na requisição, além de determinar o mês e o ano dos dados. Novamente para o caso de não carregamento dos dados, o tratamento utilizado foi o mesmo visto na Seções 5.1.3 e 5.1.2.

Já em relação a questão de armazenamento de dados (ver Figura 4), novamente foi usado o *BigQuery*. O tratamento de erros também foi salvo no conjunto de dados *err* na tabela *leitos\_uti*. Quando o erro não acontece, é armazenado no conjunto *dados\_extras* na tabela, também chamada de, *leitos\_uti*



Figura 4 – Dados Extras - BigQuery

#### 5.1.5 SET Datasets

Para os conjuntos de dados que não passaram por nenhum processo automático, ou seja, sua extração foi realizada por meio de um *download* de um arquivo CSV, foi necessário que eles passassem por algumas modificações. Logo, são os conjuntos relacionados a casos e óbitos de SRAG, a ocupação dos leitos hospitalares e aos respiradores.

Como nesses *datasets* foram apenas realizados o *download*, muitas informações desnecessárias vieram com eles, por isso foi utilizado o *Google Colaboratory*. Assim, foram aplicados alguns filtros, para diminuir a dimensionalidade dos dados, selecionando apenas algumas colunas, as datas passaram por um filtro e além de algumas outras modificações nos dados necessárias para melhor realizar as análises.

## 5.2 Pré-Análise

Esta seção irá descrever dois pontos necessários para realizar as análise são eles o carregamento de dados e o pré-processamentos dos mesmos, chamado de *Data Munging*.

### 5.2.1 GET Data

Essa subseção irá discorrer sobre o processo de carregamento dos dados para o *script*. Foi criado um objeto global para cada conjunto de dados, ficando da seguinte forma:

1. **Conjunto IDH** chamado de *idh\_*, composto por um *dataframe*
2. **Conjunto Leitos de UTI** chamado de *leitos\_uti\_*, composto por um *dataframe*
3. **Conjunto Casos e Óbitos SRAG** chamado de *casos\_obitos\_*, composto por um *dataframe*
4. **Conjunto Ocupação Leitos** chamado de *ocup\_leitos\_*, composto por um *dataframe*
5. **Conjunto Respiradores** chamado de *respiradores\_*, composto por um *dataframe*
6. **Conjunto Ações Orçamentárias** chamado de *acao\_*, composto por um dicionário, onde cada chave é um identificador de ação e seu respectivo valor corresponde aos dados, outro dicionário, onde cada chave é a tabela e o valor um *dataframe*. (Veja Código 5.5)

```
acao_ = {
    key_id_acao1 : {
        level01: dataframe_object,
        key_programa_id1: dataframe_object,
        key_programa_id2: dataframe_object
    },
    key_id_acao2 : {
        level01: dataframe_object,
        key_programa_id: dataframe_object
    }
}
```

Código 5.5 – Arquitetura Objeto acao\_

Note que cada objeto possui o sufixo *\_*, para assim ser identificado com o objeto original. Quando as análises forem feitas, será feita uma cópia desses dados, com o objetivo de manter os dados originais.

Para ler os dados, foram criadas duas funções auxiliares (Ver Código 5.6). A função *read\_bq()* irá receber nome o conjunto de dados (variável *dataset*) e a nome da tabela (variável *table*) e irá retornar o *dataframe* lido por meio de uma *query* no *Big Query*. Já a função *read\_csv()* receberá o nome do arquivo (variável *table*) e retornará o *dataframe* lido de um arquivo CSV.

```
def read_bq(dataset, table):
    credentials = get_credentials()
```

```

query = f"""select * from 'tcc-308813.{dataset}.{table}';"""
df = pd.read_gbq(query, credentials=credentials)
return df

def read_csv(table):
df = pd.read_csv(f'/content/drive/MyDrive/TCC/dataset/{table}.csv')
return df

```

Código 5.6 – Leitura dos Dados

## 5.2.2 Data Munging

Nesta seção, os conjuntos de dados vão ser passados por algumas modificações, o intuito delas é manter um padrão nos dados, como os nomes dos estados, além de mudar nomenclaturas de colunas (todos os objetos tiveram algumas colunas renomeadas) e tipos de dados. Para todos os conjuntos, serão mostradas algumas informações, oriundas da função `show_info()` (Veja Código 5.7), serão elas: a quantidade de linhas e colunas, o nome da colunas, a tipagem dos dados, e um resumo estatístico para as variáveis categóricas e numéricas (se existirem).

```

def show_info(df):
    print("Numeros de Linhas: ", df.shape[0])
    print("Numeros de Colunas: ", df.shape[1])
    print("Colunas: ", df.columns.tolist())
    print('\n-----\n')
    print("Tipos de Dados: ")
    print(df.info())
    print('\n-----\n')
    summary = df.describe(include='object').T
    summary['null values'] = df.isnull().sum()
    print("Summary (tipo objetos):")
    print(summary)
    print('\n-----\n')
    try:
        summary = df.describe(include=['int', 'float']).T
        summary['null values'] = df.isnull().sum()
        print("Summary (tipo numerico):")
        print(summary.iloc[:, :4])
        print(summary.iloc[:, 4:])
        print('\n-----\n')
    except ValueError:
        pass
    print('Primeiras 5 linhas:')

```

```
print(df.head())
```

Código 5.7 – Função show\_info()

O objeto `idh_` trata os estados colocando todos com o nome por extenso, com a primeira letra maiúscula e sem acentos (esse será o padrão utilizados para os estados). Em `leitos_uti_`, os acentos são retirados da colunas dos estados (mantendo o padrão) e é extraído o número do mês de cada amostra, enquanto que o objeto `ocup_leitos_` os acentos dos estados são retirados e é extraído o número de mês e do dia. Já o `casos_obitos_` e `respiradores_` também modificam a coluna de estado e extraem da coluna `data`, o número do mês e do dia, sendo que para o primeiro, há a mudança de alguns tipos de dados. Por último, no objeto `acao_` é tratada a questão da nomenclatura dos estados, são extraídos o mês e dia (se existir), além de algumas alterações em colunas que serão passadas para o tipo numérico e também foi criado um dicionário com os identificadores e nome das ações e programas.

## 5.3 Análises

Nessa seção serão apresentados todos os códigos utilizados para realizar a análise de dados.

### 5.3.1 Núcleo 1

Esse núcleo busca entender qual era a situação em que as Unidade Federativas do Brasil estavam, ou seja, serão analisados todos os conjuntos de dados (menos o das ações), para melhor entendermos qual era a situação de cada estado em determinado pontos da pandemia durante 2020, de Fevereiro até Dezembro.

#### 5.3.1.1 Índice de Desenvolvimento Humano - IDH

Na Figura 5, é apresentado o gráfico do IDH de cada Unidade Federativa do Brasil. IDH é o acrônimo de Índice de Desenvolvimento Humano, ou seja, é um índice que leva em conta a saúde, educação e renda que uma população de um determinado território possui. Ele foi inventado pela ONU - Organização das Nações Unidas, como uma forma de medir o avanço dos seus países-membros.

Com base nele, é possível concluir que, apesar do país possuir, na média, o IDH um pouco superior a 0,7, sendo classificado como Alto Desenvolvimento Humano, a maioria dos estados estão abaixo dessa linha, mostrando que apenas alguns deles, como o Distrito Federal, Santa Catarina e São Paulo, elevam o índice para cima, uma visão preocupante, uma vez que há muitos abaixo de 0,65. O objetivo de conjunto é servir como uma métrica de avaliação dos estados nas análises posteriores.

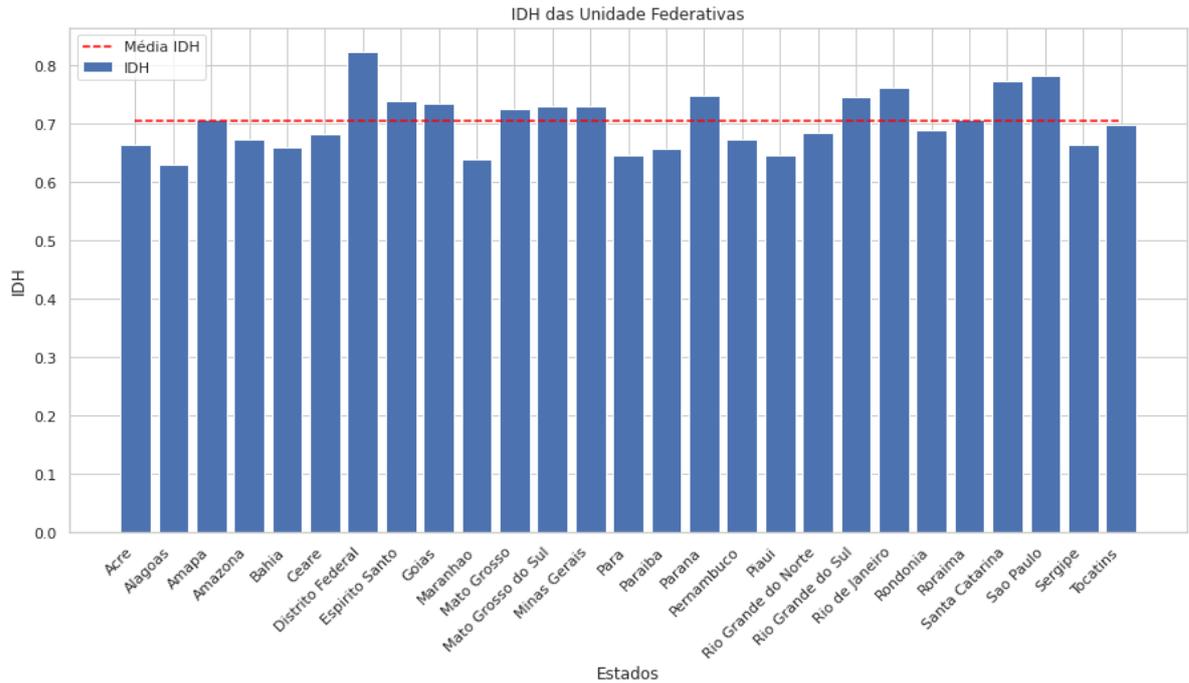


Figura 5 – IDH das Unidades Federativas

### 5.3.1.2 Casos e Óbitos de SRAG

O próximo conjunto é os dos casos e óbitos e, por meio de vários histogramas (veja Figura 6), será avaliada a distribuição dos dados. O perfil dos doentes é homem (cerca de 60% das vezes) branco ou pardo, entre 50 e 70 anos. O nível de escolaridade não pode ser medido com precisão, pois a maioria dos dados ou estão nulos ou foi ignorado.

Como já é sabido, esse conjunto de dados é sobre as ocorrências de SRAG, que pode ser adquirido pelo *Influenza*, ou outro vírus respiratório, ou agente etiológico ou então do COVID-19. Consegue-se confirmar que todos esses casos referem-se a pacientes que foram infectados pelo *SARS-CoV-2* por meio do histograma dos diagnósticos. O último mostra que a maioria dos pacientes se recuperam, porém, como a quantidade de infectados é alta, logo a quantidade de óbitos também será.

O resultado final do tratamento de uma doença é uma métrica importante. A Figura 7 foca na evolução comparando com o sexo e a raça/cor dos participantes. Olhando apenas para o sexo, percebe-se que a quantidade de altas e óbitos para homens é maior que para mulheres, o que faz sentido, uma vez que o número de pacientes contaminados é maior entre os homens. Agora, quando olhamos a raça/cor, a quantidade de altas faz sentido, pois a maioria dos casos estão entre os brancos e pardos, respectivamente, porém, a quantidade de óbitos foge do esperado, pois, entre brancos e pardos, é a mesma, apesar da quantidade de casos em brancos ser maior.

Focando para os dados de casos ao decorrer do ano, a Figura 8 mostra um gráfico de linhas contendo os seis estados que mais tiveram ocorrência da doença. A partir disso,

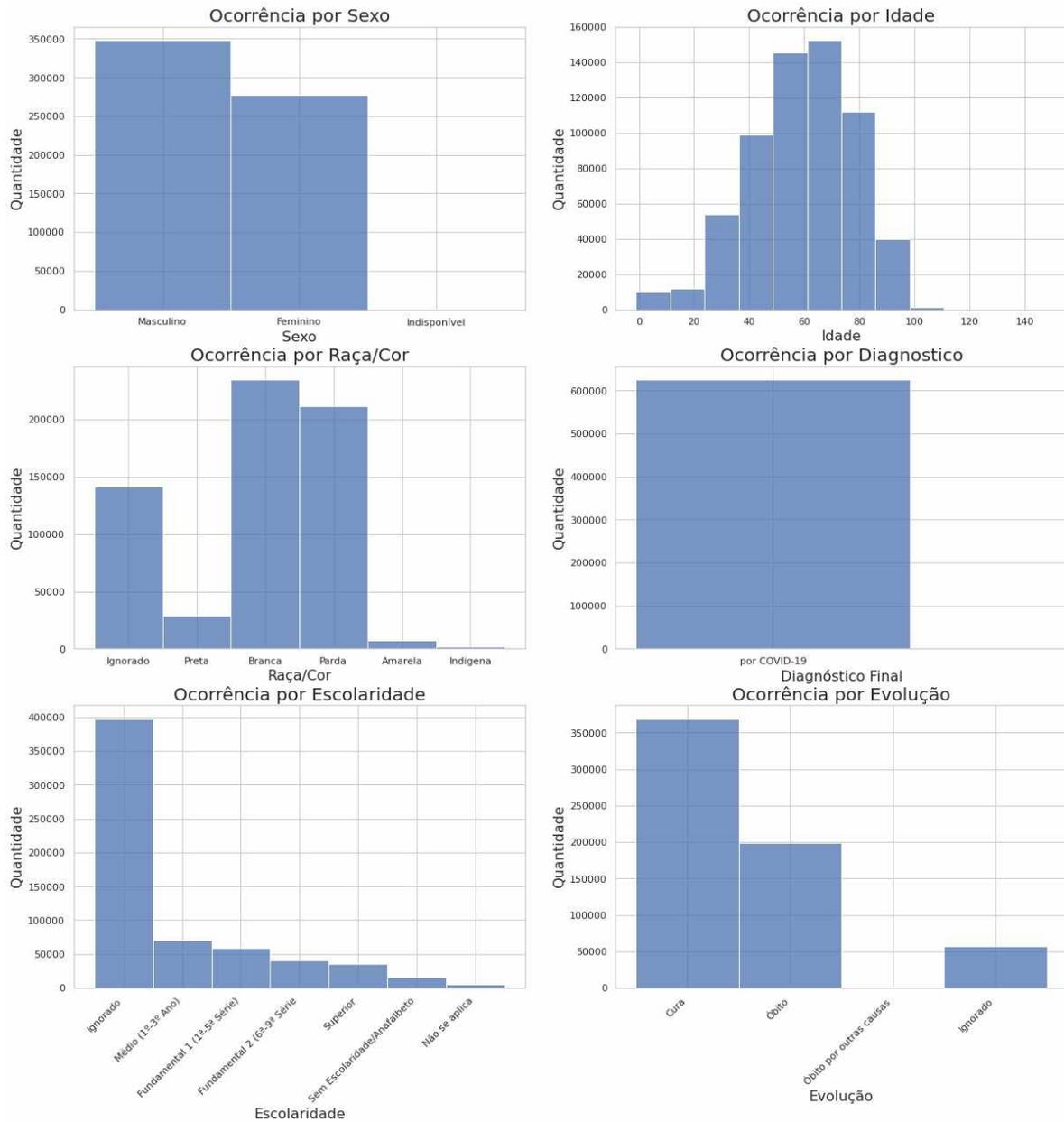


Figura 6 – Histogramas dos Casos de SRAG

consegue-se perceber padrões, como a alta que os estados de São Paulo, Rio de Janeiro e Pernambuco tiveram juntos no mês 03 e 04, e a queda no mês 05, o que era esperado, pois são estados populosos e que atraem muitos turistas, internacionais e nacionais. Já Minas Gerais, Paraná e Rio Grande do Sul tiveram altas e quedas bem parecidas, no início de forma diferente dos outros três estados, mas, no final do ano, obtiveram altas parecidas. Esses seis serão os estados da análise.

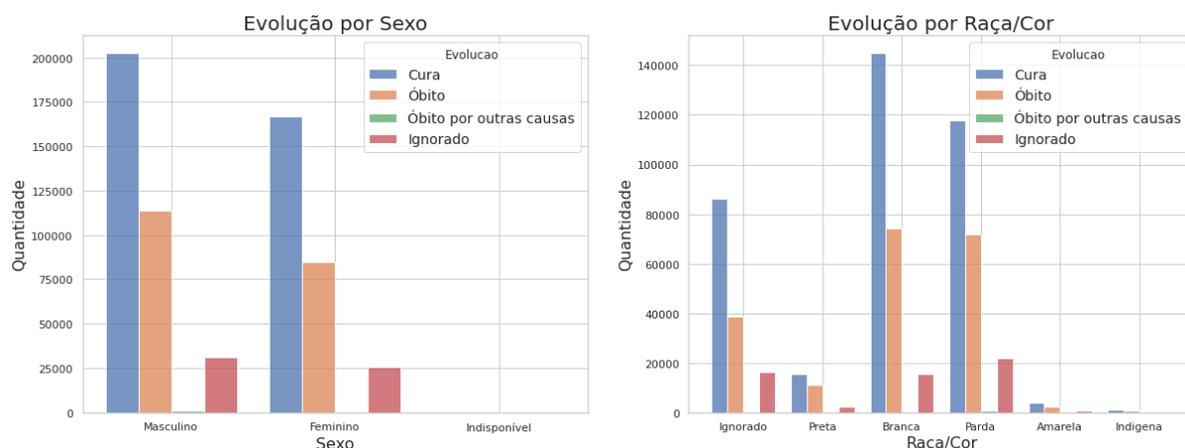


Figura 7 – Evolução das ocorrências por Sexo e Raça

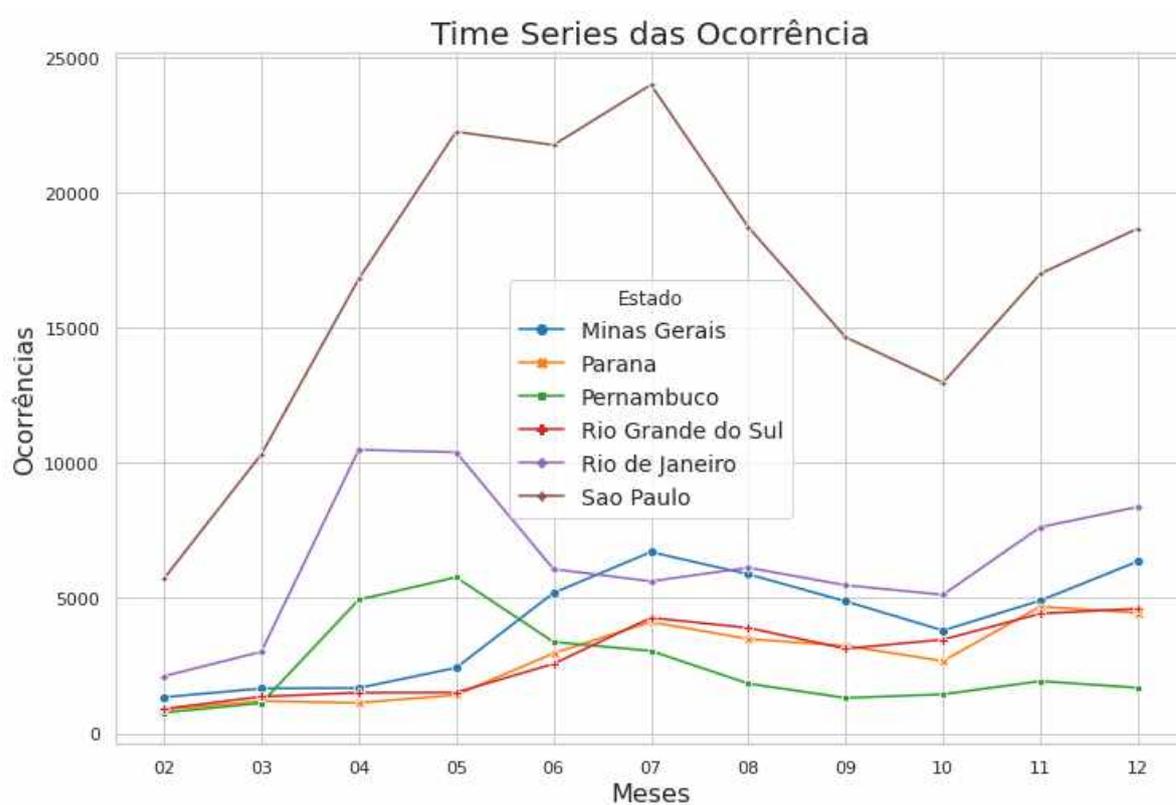


Figura 8 – Ocorrência por estados

### 5.3.1.3 Quantidade de Leitos de UTI

A Figura 9 mostra a quantidade de leitos de UTI distribuídos pelo país. Novamente, iremos selecionar os mesmos seis estados: Minas Gerais, Paraná, Pernambuco, Rio Grande do Sul, Rio de Janeiro e São Paulo. A primeira informação que obtemos por meio desse gráfico é que, a partir do mês 03, todos passaram a ter leitos para COVID, o que era esperado, pois a pandemia começou no mês anterior, em Fevereiro. Três estados se des-

tacam pela alta na quantidade de leitos, sendo eles, por ordem do maior para o menor, São Paulo, Rio de Janeiro e Minas Gerais.

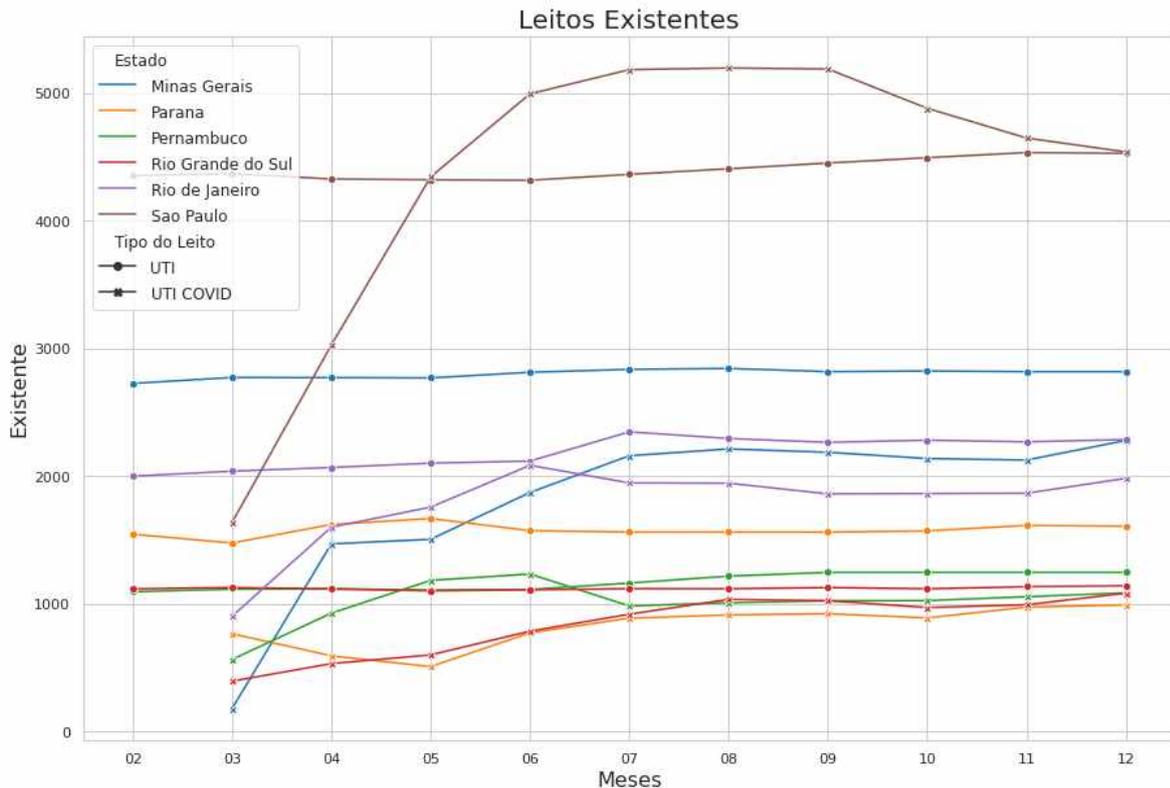


Figura 9 – Distribuição dos Leitos de UTI

#### 5.3.1.4 Ocupação dos Leitos Hospitalares

Esse conjunto refere-se à ocupação dos leitos hospitalares durante a pandemia, referentes aos seis estados selecionados. Os dois gráficos da Figura 10 representam a quantidade de leitos, à esquerda são clínicos e à direita são UTIs, ocupados por cada estado. Em ambos os gráficos, segue-se uma tendência de alta, principalmente São Paulo, onde houve uma explosão de ocupação, plausível com o que já foi visto.

#### 5.3.1.5 Respiradores

Os gráficos da Figura 11 apontam os principais fornecedores de respiradores, levando em conta o valor pago pelos respiradores, sendo representado pelo gráfico à esquerda, e pela quantidade comprada, gráfico à direita. Vale destacar as empresas *Vyair* e *Magnamed*, ambas são as líderes isoladas na quantidade vendida de respiradores e o preço pago por eles, onde a primeira é a líder entre os Respiradores de UTI e a segunda a líder entre os de transporte.

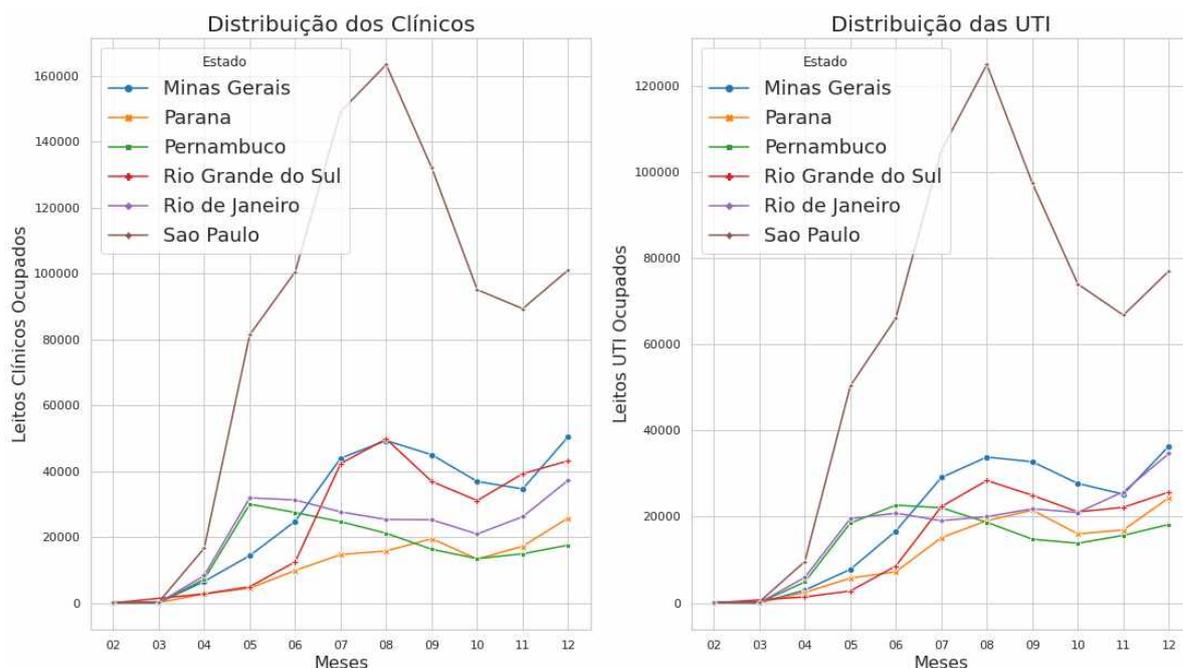


Figura 10 – Ocupação dos Leitos Hospitalares

No gráfico da Figura 12, é mostrada a situação de seis estados da análise, em relação aos respiradores. Quando comparado um gráfico com o outro (o de valor com o de quantidade), percebe-se que os picos são parecidos e proporcionais. Destacam-se os estados de São Paulo, Rio de Janeiro e Rio Grande do Sul, com os maiores picos, principalmente os dois primeiros, onde o pico de compras foi logo após seguidas altas de casos.

### 5.3.2 Núcleo 2

Os gráficos da Figura 13 representam os gastos de cada uma das Ações Orçamentárias. Há dois gráficos, onde irão apresentar os valores gastos, sendo, o Valor Pago, que é a quantidade de verba já recebida e, o Valor Faltante, que representa o valor que falta ser enviado.

Do primeiro, é possível observar as quatro ações que mais receberam verba, sendo, do maior para o menor, 00S4 - Auxílio Emergencial de Proteção Social a pessoas de vulnerabilidade, 00EE - Integralização de cotas no Fundo Garantidor de Operações (FGO) para o Programa de Apoio as Microempresas e Empresas de pequeno porte, 00S7 - Auxílio financeiro aos estados, ao Distrito Federal e aos municípios e, em quarto, 00SF - Auxílio emergencial residual para enfrentamento da emergência de saúde pública. As ações 00ED - Integralização de cotas do Fundo Garantidor de Crédito para micro, pequenas e médias empresas para o Programa Emergencial de Acesso a Crédito e 00SG - Aporte para agente financeiro BNDES para concessão de empréstimos no âmbito do Programa Emergencial de Acesso a Crédito PEAC, ficam empatadas em quinto.

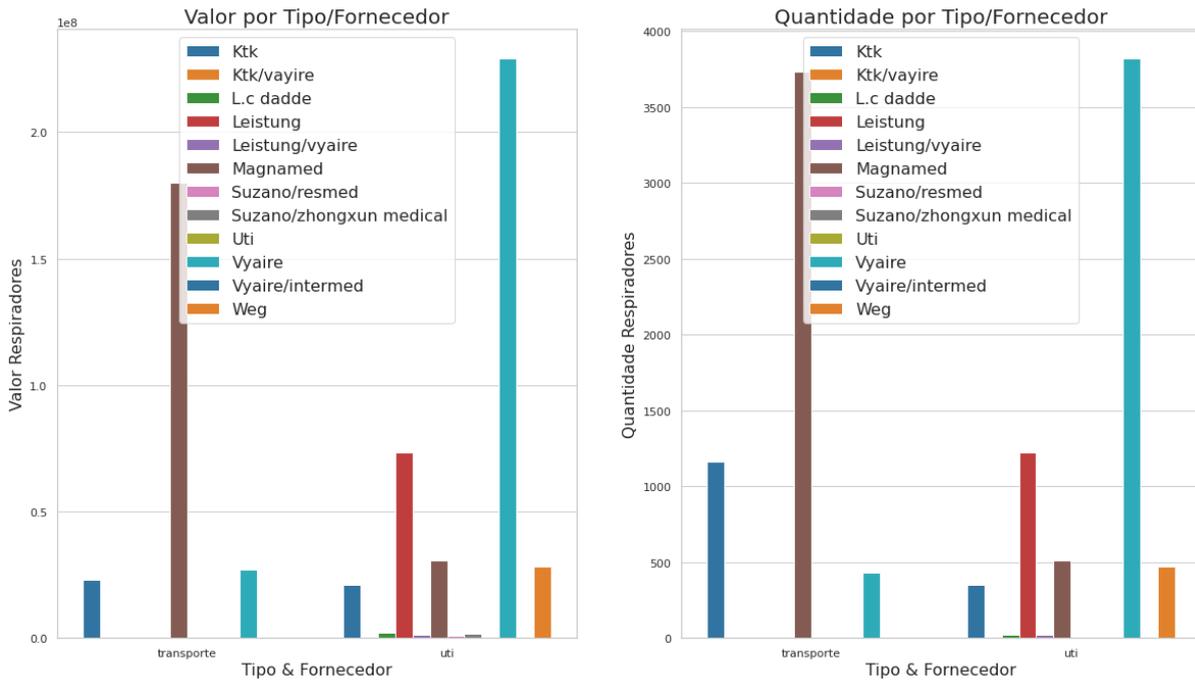


Figura 11 – Fornecedores de Respiradores

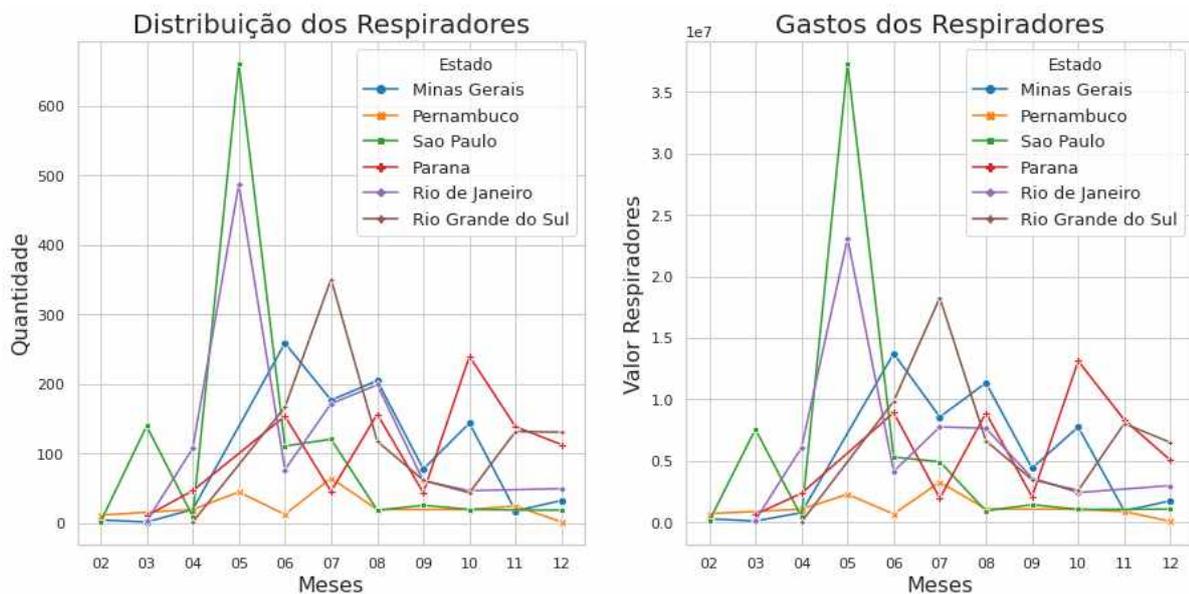


Figura 12 – Gastos e Quantidade de Respiradores

Das seis ações citadas anteriormente, apenas a 00SF falta um montante relativamente grande para ser depositado. Além disso, no segundo gráfico também foi possível verificar os maiores montantes que ainda não foram liberados, sendo a 20RK - Funcionamento de Instituições Federais de ensino superior, 0515 - Dinheiro direto na escola para a educação básica e 2E89 - Incremento temporário ao custeio dos serviços de atenção primária à

saúde, as três principais ações que mais não receberam o dinheiro.

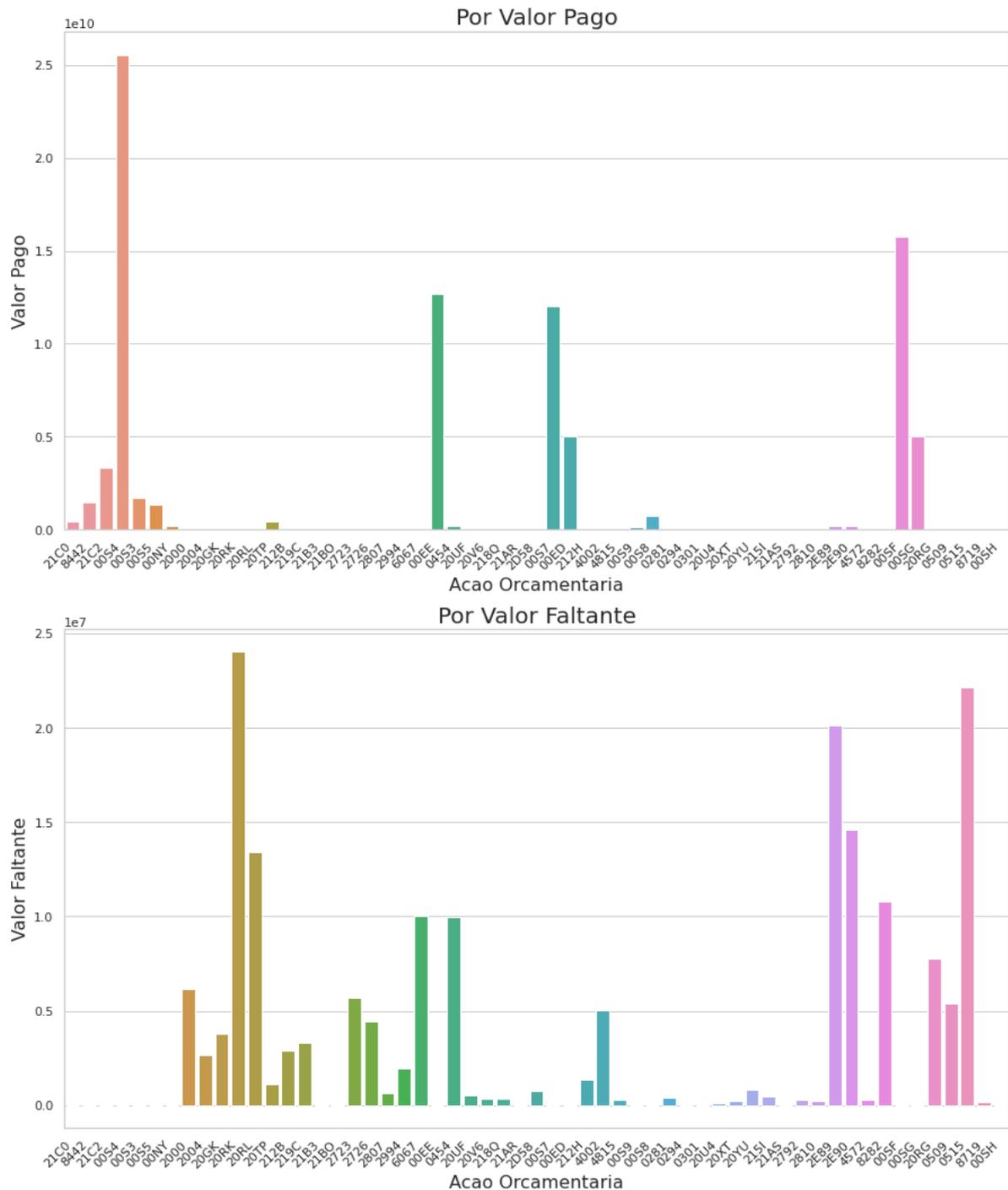


Figura 13 – Valor Pago e Faltante das Ações Orçamentárias

Os gráficos da Figura 14 irão apresentar alguns aspectos das ações, foi utilizado o conjunto de dados contendo mais detalhes (são as tabelas terminadas com \_detail), deve-se notar que os valores estão na ordem de grandeza  $10^5$ . Foram selecionadas as quatro maiores Ações Orçamentárias, levando em consideração a verba de cada uma, são elas:

1. **00ED:** Integralização de Cotas do Fundo Garantidor de Crédito para micro, pequena e médias empresas para o Programa Emergencial de Acesso a Crédito;
2. **00SG:** Aporte para agente financeiro BNDES para concessão de empréstimos no âmbito do Programa Emergencial de Acesso a Crédito PEAC;
3. **21C0:** Enfrentamento da emergência de saúde pública de importância internacional decorrente do Coronavírus;
4. **00NY:** Transferência de recursos para a conta de Desenvolvimento Energético.

O gráfico da esquerda irá apresentar os Órgãos que mais receberam verba, sendo o Ministério da Economia, de forma disparada, o maior deles, seguido pelo Ministério de Minas e Energia e o Ministério da Saúde. O auxílio emergencial foi disponibilizado por meio do Ministério da Economia, o que explica a diferença gritante entre os órgãos. Já no gráfico à direita, mostra que a maioria da verba está na fase de Empenho, que significa que o estado reservou essa verba, e na fase de Pagamento, que representa quando o pagamento é efetuado. A fase de Liquidação, que é quando se verifica a origem e o objeto que se deve pagar, quase não há montante destinado.

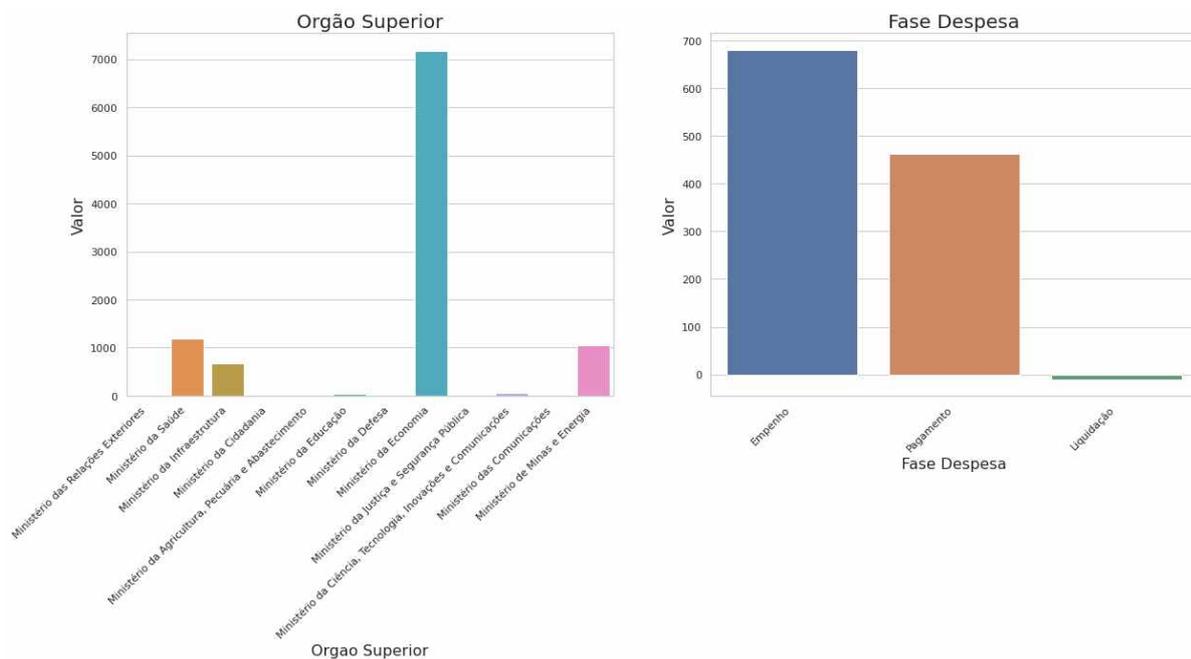


Figura 14 – Gastos por Órgão Superior e Fase da Despesa

### 5.3.3 Núcleo 3

Os gráficos das Figuras 15 e 16 representam os dados de todos os conjuntos de dados analisados no Núcleo 1 e a Tabela 1, o gasto total das Ações Orçamentárias, as quatro

maiores já faladas anteriormente. Os valores estão no ordem de grandeza  $10^5$ . Ambas informações são referentes aos seis estados da análise: Minas Gerais (MG), Paraná (PR), Pernambuco (PE), Rio Grande do Sul (RS), Rio de Janeiro (RJ) e São Paulo (SP), no decorrer do ano de 2020. Será realizada uma análise mais detalhada para cada estado.

Todos os seis estados selecionados para análise possuem IDH superior a 0,7 (menos o estado de Pernambuco) e foram os estados que mais sofreram com altos casos de COVID durante 2020. Em relação à quantidade de Leitos de UTI, todos os estados sofreram um aumento no início do ano, principalmente o estado de São Paulo, porém, por volta dos meses 06 e 07, houve uma estagnação na alta.

**Minas Gerais:** Se olharmos para a ocupação dos leitos, tanto de UTI quanto clínicos, há uma crescente iniciando no mês 05 e atingindo o pico no mês 08. Esse movimento condiz com a quantidade de casos, que iniciou nos meses 04 e 05 e atingiu o pico em 07, as curvas de óbitos e alta seguiram a mesma tendência. Os maiores picos de compra de respiradores foram durante os meses 06, 07 e 08, durante a alta de ocupação de leitos. O estado, desde o mês 03 recebe quantias significativas, sempre superiores a R\$ 30 milhões. Destaque para os meses 05, 07 e 08, que receberam mais de 145 milhões e são os meses que o estado estava com a situação complicada.

**Paraná:** A ocupação do leitos clínicos não sofreu grande subida, apesar de ter sofrido uma leve alta, tendo o pico no mês 09, mesmo mês em que a curva dos leitos de UTI teve pico, porém eles tiveram uma alta maior que os clínicos. Olhando pros casos, o seu pico foi no mês 07 e 08, gerando os picos nos leitos, as curvas dos óbitos e alta seguem a mesma tendência. Interessante observar uma alta de casos no mês 11, porém, como os dados dos leitos vão só até dezembro de 2020, não poderá ser verificado se houve um pico nas ocupações no início de 2021. Em vez de um grande pico na compra dos respiradores, o estado comprou, a cada dois meses, começando no mês 06 e mantendo alta nos meses 11 e 12, quando também houve alta de casos. Sobre as ações, o estado se destaca nos meses 05, 07 e 11, sendo o último o maior deles.

**Pernambuco:** O estado sofreu grande alta na ocupações dos leitos, tanto clínicos quanto UTI, no início da pandemia, principalmente a partir do mês 05, o que coincide com a alta dos casos que o estado sofreu no início do ano, meses 03 e 04. As curvas de altas e óbitos seguem a mesma tendência, com destaque para a segunda, que, durante os meses 04 e 05, teve uma grande alta. A compra dos respiradores não foi tão grande, comparando com os outros estados, tendo um pico no mês 07. Em relação às Ações, os meses 06 e 07 foram os maiores, que são os meses posteriores aos que houve pico, mostrando uma resposta da União frente ao COVID.

**Rio Grande do Sul:** A ocupação dos leitos clínicos e UTI sofreu grande alta no mês 06, atingindo o pico no mês 08 e tendo uma baixa até o mês 10 e voltando a subir. A quantidade de casos teve o pico no mês 07 e depois no mês 11, e as curvas de óbitos e altas seguindo a mesma tendência. Nesse mês, 07, também houve uma grande compra

de respiradores e o segundo pico sendo no mês 11. Três meses se destacam entre os valores das ações, sendo eles o mês 05, 07 e 12, no qual os dois últimos foram meses que reportaram aumento de ocorrência.

**Rio de Janeiro:** Como Pernambuco, o estado sofreu um aumento na ocupação dos Leitos clínicos e UTI, e a partir do mês 05, coincidindo também com a alta de casos, ocorridas no mês 04, porém, apesar da tendência ser a mesma, a quantidade de casos é superior no Rio. As curvas de altas e óbitos seguindo a mesma tendência, porém com uma alta expressiva na última, durando os meses 04 e 05. A compra de respiradores obteve seu pico no mês 05, no segundo maior pico de mortos, um mês depois do maior pico. Sobre as Ações, o estado foi um dos principais auxiliados, tendo, nos meses 06, 08, 09 e 10, recebido valores que ultrapassam a casa do trilhão de real.

**São Paulo:** Como os estados do Rio e Pernambuco, São Paulo sofreu muito no início, tendo a ocupação de leitos clínicos e UTI crescendo a partir do mês 04 e atingindo a alta no mês 08, ambos os pontos aconteceram, respectivamente, um mês após iniciar a alta nos casos e seu pico. Esse período de alta foi maior no tempo e na quantidade de casos do que foi nos outros estados. As curvas de altas e óbitos seguiram a mesma tendência. Os respiradores sofreram um pico de compra durante o mês 05, bem maior que qualquer outro pico de outro estado analisado, e depois foi diminuindo ao longo do ano. Como o estado é um dos mais populosos do Brasil e foi o que mais sofreu com os casos, tirando o mês 02, todos os meses foram liberados mais de R\$ 10 milhões, com destaque para o mês 04 e 08, que ultrapassaram o bilhão, sendo, respectivamente, 7 e 4 bilhões de reais.

|            | MG           | PR         | PE       | RJ          | RS       | SP         |
|------------|--------------|------------|----------|-------------|----------|------------|
| <b>Mês</b> | <b>Valor</b> |            |          |             |          |            |
| 02         | 0            | 0          | 0        | 0           | 0        | 875,7330   |
| 03         | 408,5795     | 67,8410    | 26,7953  | 114,5033    | 0        | 1400,6120  |
| 04         | 808,0855     | 36706,9009 | 149,0380 | 1361,8283   | 42,1170  | 70416,7600 |
| 05         | 1928,3601    | 243,7979   | 0        | 769,5211    | 276,9551 | 9008,0633  |
| 06         | 437,3095     | 15,3476    | 508,5069 | 50510,5648  | 31,5488  | 4528,4521  |
| 07         | 1493,9501    | 210,7652   | 389,7207 | 346,983     | 225,4779 | 4313,3991  |
| 08         | 1960,9678    | 64,3272    | 47,9751  | 100197,9518 | 77,5462  | 46107,7573 |
| 09         | 884,5392     | 36,8685    | 0        | 360869,8742 | 143,0913 | 1968,6563  |
| 10         | 985,2806     | 142,2020   | 506,9115 | 104406,4582 | 48,0109  | 1124,3045  |
| 11         | 374,3840     | 1332,9909  | 0,3042   | 5733,3595   | 65,7843  | 4286,8540  |
| 12         | 435,5926     | 131,8044   | 230,6023 | 9399,2771   | 247,7656 | 7832,2784  |

Tabela 1 – Gastos das Ações Orçamentárias por Estado

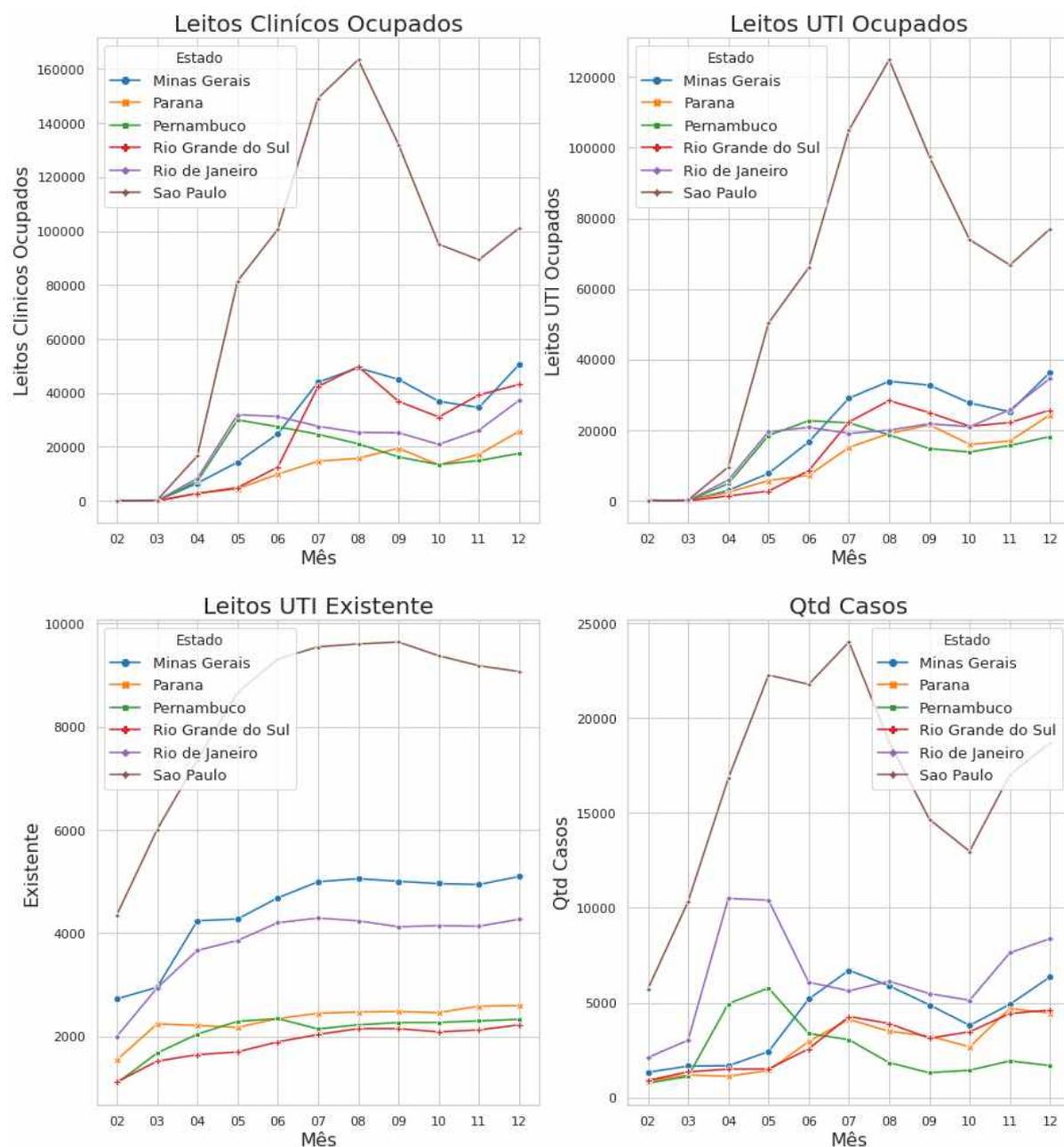


Figura 15 – Visão Geral da Situação do Brasil - Parte 01

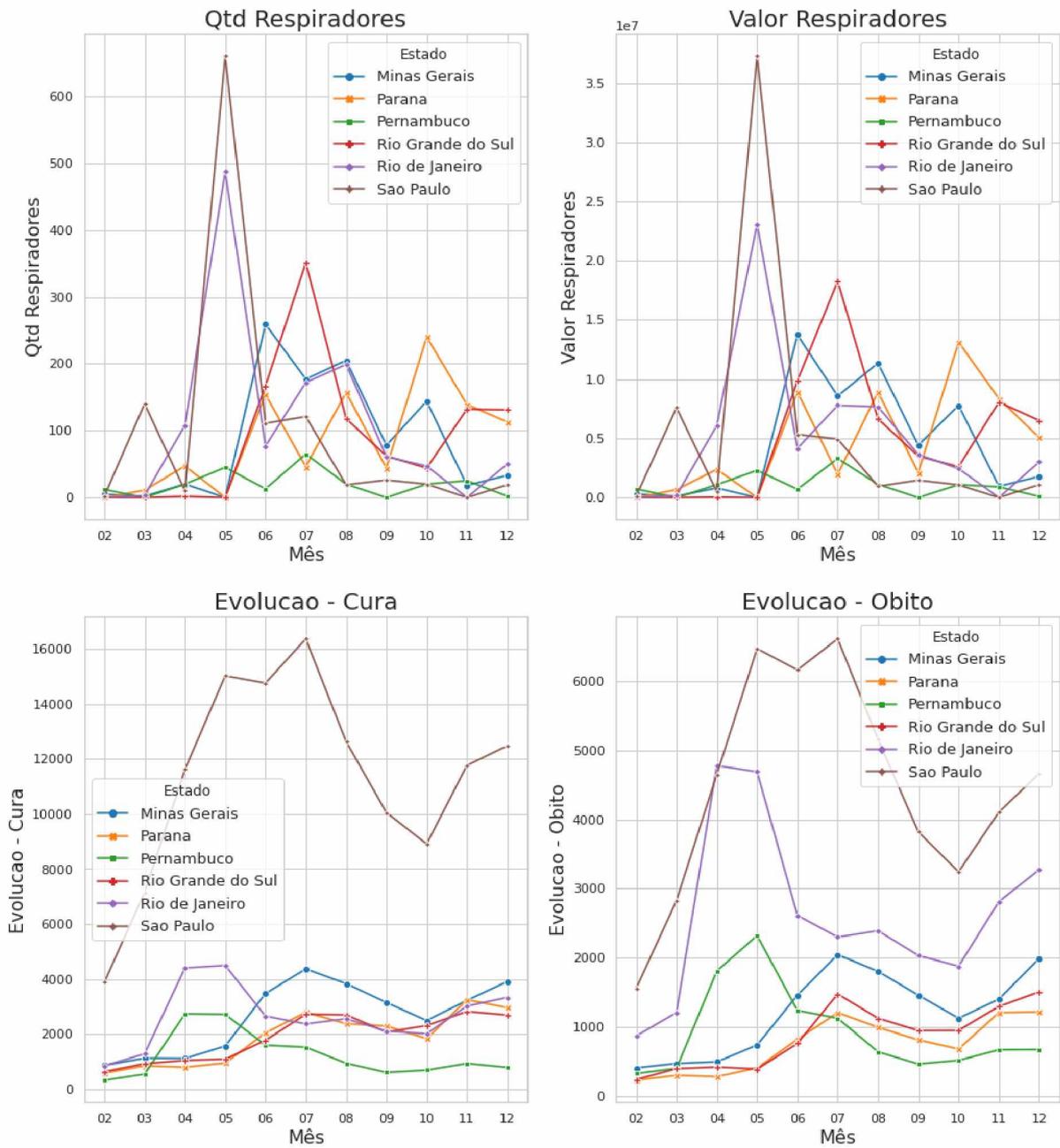


Figura 16 – Visão Geral da Situação do Brasil - Parte 02

---

## Conclusão

Neste capítulo, serão apresentadas as conclusões obtidas por meio das análises anteriores, além de, em um segundo momento, discorrer sobre possíveis trabalhos que podem ser criados a partir deste.

### 6.1 Avaliação dos resultados

No núcleo 1, foi obtida a visão geral das unidades federativas do Brasil durante os meses de Fevereiro até Dezembro de 2020 para alguns atributos, como a quantidade de leitos hospitalares, casos de COVID-19 e outras métricas relacionadas à saúde. No núcleo 2, foi a vez de serem visualizadas algumas características das Ações Orçamentarias, como a fase de pagamento, quantidade da verba destinada aos órgãos superiores, o valor pago e faltante.

O núcleo 3 é a junção das análises do núcleo 1 e 2, porém de forma mais aprofundado, focando em apenas alguns estados (Minas Gerais, Paraná, Pernambuco, Rio de Janeiro, Rio Grande do Sul e São Paulo) e em algumas ações (00ED, 00SG, 21C0 e 00NY). Com essa análise, foi percebida a resposta rápida que a União obteve em praticamente todos os estados analisados no momento em que o número de casos crescia ou então logo no mês seguinte a um aumento de casos.

Em relação a outros fatores, a quantidade de leitos hospitalares sofreu grande aumento no primeiro semestre, porém se estagnaram. Os respiradores tiveram picos de compras, principalmente no primeiro semestre, e, apesar de terem diminuído, a compra continuou em níveis menores. A ocupação dos leitos hospitalares acompanha a curva de casos, assim como as de altas e óbitos.

Dessa forma, foi visto que a União agiu nos momentos necessários, apesar de alguns casos com atrasos, auxiliando os estados a melhor combater a pandemia. Importante lembrar que os dados referentes às ações só garante que o Governo Federal liberou a verba para uma determinada instituição de um determinado estado. Para verificar se o dinheiro foi corretamente gasto, é necessária uma análise mais focada.

## 6.2 Avaliação futuras

Como dito anteriormente, alguns questionamentos ficaram sem resposta, isso porque, para responder alguns indagações, necessita-se de uma análise mais focada naquele assunto. Como já dito, para verificar se a verba recebida pelos estados foi devidamente gasta e não houve desvio ou qualquer outra irregularidade, é necessário analisar como e com o que aquela unidade federativa gastou o dinheiro recebido.

Outros questionamentos surgiram, como em relação aos respiradores. Na Figura 11, foi visto que alguns fornecedores receberam mais por cada aparelho vendido, vale analisar se aquele montante é coerente ou houve alguma irregularidade. A questão da estagnação dos leitos hospitalares a partir dos meses 06 e 07 pode ser melhor analisada.

O Brasil é composto por vinte e sete Unidades Federativas e foram divulgadas mais de cinquenta Ações Orçamentárias. Com base nisso, a análise não abrangeu todas as possibilidades e cenários, além de focar apenas nos meses de Fevereiro a Dezembro de 2020. Das vinte e sete Unidades Federativas, apenas seis foram analisadas, e das ações, apenas quatro foram levadas em conta.

Logo, analisar cada ação e se seu propósito foi devidamente cumprido, como se os outros estados também receberam auxílio quando necessário, é interessante mas precisará de mais aprofundamento e possivelmente mais bases de dados sobre outras características relacionadas à pandemia.

---

## Referências

ALBERTINI, M. K.; BACKES, A. R.; SÁ, A. L. A study of publication trajectories of the brazilian computer science community. **Anais da Academia Brasileira de Ciências**, SciELO Brasil, v. 91, n. 3, 2019.

ALURA. **Google Colab: o que é e como usar?** 2021. Disponível em: <<https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar>>.

APARICIO, M.; COSTA, C. J. Data visualization. **Communication design quarterly review**, ACM New York, NY, USA, v. 3, n. 1, p. 7–11, 2015.

BCCNEWS. **Coronavirus confirmed as pandemic by World Health Organization**. 2021. Disponível em: <<https://www.bbc.com/news/world-51839944>>.

\_\_\_\_\_. **Origem do coronavírus: de morcegos a laboratório, veja as conclusões da investigação da OMS na China**. 2021. Disponível em: <<https://www.bbc.com/portuguese/geral-56587394>>.

BEHRENS, J. T. Principles and procedures of exploratory data analysis. **Psychological Methods**, American Psychological Association, v. 2, n. 2, p. 131, 1997.

CLOUDXLAB. **NumPy and Pandas Tutorial - Data Analysis with PYthon**. 2021. Disponível em: <<https://cloudxlab.com/blog/numpy-pandas-introduction/>>.

COOK, D.; SWAYNE, D. F. **Interactive and Dynamic Graphic for Data Analysis With R and GGobi**. [S.l.]: Springer Verlag, 2007.

ECONOMIA, M. da. **Monitoramento dos Gastos da União com Combate à COVID-19**. 2021. Disponível em: <<https://www.tesourotransparente.gov.br/>>.

ENDEL, F.; PIRINGER, H. Data wrangling: Making data useful again. **IFAC-PapersOnLine**, Elsevier, v. 48, n. 1, p. 111–112, 2015.

ESTATÍSTICA, . I. I. B. de Geografia e **IBGE - Instituto Brasileiro de Geografia e Estatística**. 2021. Disponível em: <<https://cidades.ibge.gov.br/>>.

FENASAUDE. **Bill Gates alerta para o risco de uma pandemia mundial matar até 30 milhões de pessoas em um ano**. 2021. Disponível em: <<https://fenasaude.org.br/noticias/bill-gates-alerta-para-o-risco-de-uma-pandemia-mundial-matar-ate-30-milhoes-de-pessoas-em-um-ano.html>>.

- FRIENDLY, M. A brief history of data visualization. In: **Handbook of data visualization**. [S.l.]: Springer, 2008. p. 15–56.
- GERSHON, N.; PAGE, W. What storytelling can do for information visualization. **Communications of the ACM**, ACM New York, NY, USA, v. 44, n. 8, p. 31–37, 2001.
- GOOGLE. **O que é o BigQuer?** 2021. Disponível em: <<https://cloud.google.com/bigquery/docs/introduction?hl=pt-br>>.
- ISTOEDINHEIRO. **Um ano do primeiro caso de covid-19: veja a evolução da doença**. 2021. Disponível em: <<https://www.istoedinheiro.com.br/um-ano-do-primeiro-caso-de-covid-19-veja-a-evolucao-da-doenca/>>.
- JUPYTER, P. **About Us**. 2021. Disponível em: <<https://jupyter.org/>>.
- LARSSON, T. **The race to the top: The real story of globalization**. [S.l.]: Cato Institute, 2001.
- LENZERINI, M. Managing data through the lens of an ontology. **AI Magazine**, v. 39, n. 2, p. 65–74, 2018.
- MEDRI, W. Análise exploratória de dados. [http://www.uel.br/pos/estatisticaeducacao/textos\\_didaticos/especializacao\\_estadistica.pdf](http://www.uel.br/pos/estatisticaeducacao/textos_didaticos/especializacao_estadistica.pdf) Acesso em, v. 15, p. 05–13, 2011.
- MICROSOFT. **Visão geral do Visual Studio**. 2021. Disponível em: <<https://docs.microsoft.com/pt-br/visualstudio/get-started/visual-studio-ide?view=vs-2019>>.
- NEVES, M. C. et al. Análise exploratória espacial de dados sócio-econômicos de são paulo. **Salvador: GIS Brasil2000**, p. 1–11, 2000.
- OLIVEIRA, A. D.; NETO, J. G.; FERREIRA, M. M. Uso da análise exploratória de dados na avaliação de modificadores químicos para determinação direta e simultânea de metais em álcool combustível por gfaas. **Eclética Química**, SciELO Brasil, v. 31, n. 1, p. 7–12, 2006.
- PERPETUO, I. H. O.; WONG, L. Atenção hospitalar por condições sensíveis à atenção ambulatorial (csaa) e as mudanças no seu padrão etário: uma análise exploratória dos dados de minas gerais. **12º Seminário sobre a Economia Mineira**, v. 29, 2006.
- POST, F. H.; NIELSON, G.; BONNEAU, G.-P. **Data visualization: The state of the art**. Springer Science & Business Media, 2002.
- SAUDE, M. da. **Cadastro Nacional de Estabelecimentos de Saúde - Consulta**. 2021. Disponível em: <<http://cnes2.datasus.gov.br/>>.
- \_\_\_\_\_. **Dados abertos do SUS**. 2021. Disponível em: <<https://opendatasus.saude.gov.br/>>.
- SAURKAR, A. V.; PATHARE, K. G.; GODE, S. A. An overview on web scraping techniques and tools. **International Journal on Future Revolution in Computer Science & Communication Engineering**, v. 4, n. 4, p. 363–367, 2018.

- SKIENA, S. S. Data munging. In: **The Data Science Design Manual**. [S.l.]: Springer, 2017. p. 57–93.
- TECNOBLOG. **O que é o Python?** 2021. Disponível em: <<https://tecnoblog.net/405640/o-que-e-python-guia-para-iniciantes/>>.
- TUKEY, J. W. The future of data analysis. **The annals of mathematical statistics**, JSTOR, v. 33, n. 1, p. 1–67, 1962.
- UNIAO, C. G. da. **Portal de Transparência**. 2021. Disponível em: <<http://www.portaltransparencia.gov.br/>>.
- WAY, S. F. et al. The misleading narrative of the canonical faculty productivity trajectory. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 114, n. 44, p. E9216–E9223, 2017.
- WICKHAM, H. A layered grammar of graphics. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 19, n. 1, p. 3–28, 2010.
- WILKINSON, L. The grammar of graphics. In: **Handbook of computational statistics**. [S.l.]: Springer, 2012. p. 375–414.